GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 02:35:01 ; Search time 58.67 Seconds
(without alignments)
8658.089 Million cell updates/sec

Title: US-09-932-678-1
Perfect score: 2068
Sequence: 1 acaqaqqctqtqqctqqaaq.....tgacatttqqqattccccat 2068

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	48.2	2.3	8797	2	US-08-723-306-6	Sequence 6, Appli
2	48.2	2.3	8797	5	PCT-US96-10041-6	Sequence 6, Appli
3	48.2	2.3	11093	2	US-08-723-306-5	Sequence 5, Appli
4	48.2	2.3	11093	5	PCT-US96-10041-5	Sequence 5, Appli
c 5	40.4	2.0	7218	1	US-08-232-463-14	Sequence 14, Appl
6	39.2	1.9	4937	1	US-08-038-682-3	Sequence 3, Appli
7	39.2	1.9	4937	1	US-08-302-832-3	Sequence 3, Appli
8	39.2	1.9	4937	2	US-08-530-198-3	Sequence 3, Appli
9	39.2	1.9	4937	2	US-08-469-880-3	Sequence 3, Appli
10	39.2	1.9	4937	2	US-08-728-470-3	Sequence 3, Appli
11	39.2	1.9	4937	2	US-08-617-697-3	Sequence 3, Appli
12	39.2	1.9	4937	4	US-08-719-641-3	Sequence 3, Appli
13	35.2	1.7	5538	2	US-08-231-193A-55	Sequence 55, Appl
14	35.2	1.7	5538	2	US-08-486-273A-55	Sequence 55, Appl
15	35.2	1.7	5538	3	US-08-940-086A-55	Sequence 55, Appl
16	35.2	1.7	5538	4	US-08-940-035A-55	Sequence 55, Appl
17	34.6	1.7	2518	4	US-09-433-699-3	Sequence 3, Appli
18	34.4	1.7	289	4	US-09-007-005-17	Sequence 17, Appl
19	34.4	1.7	289	4	US-09-244-796-17	Sequence 17, Appl
20	34.4	1.7	1307	4	US-09-282-305-17	Sequence 17, Appl
21	34.4	1.7	9323	1	US-08-038-682-6	Sequence 6, Appli
22	34.4	1.7	9323	1	US-08-302-832-6	Sequence 6, Appli
23	34.4	1.7	9323	2	US-08-530-198-6	Sequence 6, Appli
24	34.4	1.7	9323	2	US-08-469-880-6	Sequence 6, Appli
25	34.4	1.7	9323	2	US-08-728-470-6	Sequence 6, Appli
26	34.4	1.7	9323	2	US-08-617-697-6	Sequence 6, Appli
27	34.4	1.7	9323	4	US-08-719-641-6	Sequence 6, Appli
28	34	1.6	1245	1	US-08-385-186-14	Sequence 14, Appl
c 29	33.4	1.6	10614	1	US-08-135-511-35	Sequence 35, Appl
c 30	33.4	1.6	10614	1	US-08-187-453-35	Sequence 35, Appl
31	33.2	1.6	1545	1	US-07-872-673B-2	Sequence 2, Appli
32	33.2	1.6	2454	1	US-07-872-673B-1	Sequence 1, Appli
c 33	32.6	1.6	19182	2	US-08-850-880-11	Sequence 11, Appl
c 34	32.6	1.6	19182	2	US-08-944-916-11	Sequence 11, Appl
c 35	32.4	1.6	2791	4	US-09-570-367C-1	Sequence 1, Appli
36	32.2	1.6	4856	4	US-09-045-360-1	Sequence 1, Appli
37	32.2	1.6	12225	2	US-08-822-445-11	Sequence 11, Appl
38	32.2	1.6	12225	4	US-09-396-540-11	Sequence 11, Appl
39	32.2	1.6	12616	2	US-08-822-445-9	Sequence 9, Appli
40	32.2	1.6	12616	4	US-09-396-540-9	Sequence 9, Appli
41	32.2	1.6	23674	4	US-09-773-816-1	Sequence 1, Appli
42	32	1.5	1202	3	US-09-058-489-29	Sequence 29, Appl
c 43	32	1.5	1248	3	US-08-910-505-3	Sequence 3, Appli
c 44	32	1.5	1251	3	US-08-910-505-1	Sequence 1, Appli
c 45	31.8	1.5	1338	3	US-09-027-064-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-723-306-6
; Sequence 6, Application US/08723306
; Patent No. 5856178
; GENERAL INFORMATION:
; APPLICANT: White PhD, Kenneth
; APPLICANT: Morrey PhD, John
; APPLICANT: Reed, William
; TITLE OF INVENTION: Cassette for Expression of Lytic
; TITLE OF INVENTION: Peptides in Mammalian Transgenic Organisms
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Trask Britt and Rossa
; STREET: P.O. Box 2550
; CITY: Salt Lake City
; STATE: Utah
; COUNTRY: USA
; ZIP: 84110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/723,306
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Sweigert PhD, Susan E
; REGISTRATION NUMBER: 36,289
; REFERENCE/DOCKET NUMBER: 2549
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 8015321922
; TELEFAX: 8015319168
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8797 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Construct comprising Bos
; DESCRIPTION: taurus beta casein 5' regulatory region plus genes encoding
; DESCRIPTION: amphipathic peptide and green fluorescent protein"
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
US-08-723-306-6

Query Match 2.3%; Score 48.2; DB 2; Length 8797;

```
  Best Local Similarity   52.2%;  Pred. No. 0.00033;
  Matches  107;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy     697 aaaaatttccattttgttcgaaaatcagaagaaacactggaatgttacgttcataacttac 756
           ||| |     || | |   |||||     ||||||    ||| | | ||| || | |||
Db    2278 AAAGAAGAGTATATTTATTAAAATTGCTCAGAACATCCAAATTTCAAGTTTATCATTTAT 2337

Qy     757 taaggaattagtgtatatttttccaacctgaaggcatgaaattctggaggcttattattgaaa 816
           | || |  | |  |    ||  | ||  |||||||| | | ||   ||||  || |
Db    2338 CTTACAATATTTCAAAAATATTAAAATAGATACATGAAATACAGAAGTAAATTAAAGAGA 2397

Qy     817 aactactcaagttgaatgtgaatgcatccccggcaggatattgaaagatgctgaagaaacag 876
           || || |  | || |     ||    ||  ||  ||  | ||  |    ||| || |
Db    2398 AAGTATTTATTTGTAAAAAAAAATTCTAGGTTGGACAGGGAGTACCAGGAAACAAAAA 2457

Qy     877 caactcaaacttgtggtggaacaga 901
           | | |||  |||| |  ||||||
Db    2458 ACAATGAAAAATGTGATCTGACAGA 2482


RESULT   2
PCT/US96/10041/6
; Sequence 6, Application PC/TUS9610041
;  GENERAL INFORMATION:
;    APPLICANT:  White PhD, Kenneth
;    APPLICANT:  Morrey PhD, John
;    APPLICANT:  Reed, William
;    TITLE OF INVENTION:  Cassette for Expression of Lytic
;    TITLE OF INVENTION:  Peptides in Mammalian Transgenic Organisms
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Trask Britt and Rossa
;      STREET:  P.O. Box 2550
;      CITY:  Salt Lake City
;      STATE:  Utah
;      COUNTRY:  USA
;      ZIP:  84110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/10041
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sweigert PhD, Susan E
;      REGISTRATION NUMBER:  36,289
;      REFERENCE/DOCKET NUMBER:  2549
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  8015321922
;      TELEFAX:  8015419168
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8797 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc  "Construct comprising Bos
;      DESCRIPTION:  taurus beta casein 5' regulatory region plus genes encoding
;      DESCRIPTION:  amphipathic peptide and green fluorescent protein"
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
PCT/US96/10041/6


  Query Match              2.3%;  Score 48.2;  DB 5;  Length 8797;
  Best Local Similarity   52.2%;  Pred. No. 0.00033;
  Matches  107;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy     697 aaaaatttccattttgttcgaaaatcagaagaaacactggaatgttacgttcataacttac 756
           ||| |     || | |   |||||     ||||||    ||| | | ||| || | |||
Db    2278 AAAGAAGAGTATATTTATTAAAATTGCTCAGAACATCCAAATTTCAAGTTTATCATTTAT 2337

Qy     757 taaggaattagtgtatatttttccaacctgaaggcatgaaattctggaggcttattattgaaa 816
           | || |  | |  |    ||  | ||  |||||||| | | ||   ||||  || |
Db    2338 CTTACAATATTTCAAAAATATTAAAATAGATACATGAAATACAGAAGTAAATTAAAGAGA 2397

Qy     817 aactactcaagttgaatgtgaatgcatccccggcaggatattgaaagatgctgaagaaacag 876
           || || |  | || |     ||    ||  ||  ||  | ||  |    ||| || |
Db    2398 AAGTATTTATTTGTAAAAAAAAATTCTAGGTTGGACAGGGAGTACCAGGAAACAAAAA 2457

Qy     877 caactcaaacttgtggtggaacaga 901
           | | |||  |||| |  ||||||
Db    2458 ACAATGAAAAATGTGATCTGACAGA 2482


RESULT   3
US-08-723-306-5
; Sequence 5, Application US/08723306
; Patent No. 5856178
;  GENERAL INFORMATION:
;    APPLICANT:  White PhD, Kenneth
;    APPLICANT:  Morrey PhD, John
;    APPLICANT:  Reed, William
;    TITLE OF INVENTION:  Cassette for Expression of Lytic
;    TITLE OF INVENTION:  Peptides in Mammalian Transgenic Organisms
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Trask Britt and Rossa
;      STREET:  P.O. Box 2550
;      CITY:  Salt Lake City
;      STATE:  Utah
;      COUNTRY:  USA
;      ZIP:  84110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/723,306
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sweigert PhD, Susan E
;      REGISTRATION NUMBER:  36,289
;      REFERENCE/DOCKET NUMBER:  2549
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  8015321922
;      TELEFAX:  8015419168
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11093 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc  "Construct comprising
;      DESCRIPTION:  portions of Bos taurus beta casein gene and genes encoding
;      DESCRIPTION:  amphipathic peptide and green fluorescent protein"
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1801..1834
;      OTHER INFORMATION:  /product= "beta casein exon 1"
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  3780..3842
;      OTHER INFORMATION:  /product= "beta casein exon 2"
;    FEATURE:
;      NAME/KEY:  TATA_signal
;      LOCATION:  1766..1773
```

```
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  4567..4590
;       OTHER INFORMATION:  /product= "portion of beta casein
;       OTHER INFORMATION:  exon 3"
;     FEATURE:
;       NAME/KEY:  5'UTR
;       LOCATION:  127..1800
;       OTHER INFORMATION:  /function= "5' flanking regulatory
;       OTHER INFORMATION:  region of bovine beta casein gene"
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  4587..5310
;       OTHER INFORMATION:  /product= "Green fluorescent
;       OTHER INFORMATION:  protein"
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  5320..5449
;       OTHER INFORMATION:  /product= "Shiva-1 coding sequence"
;     FEATURE:
;       NAME/KEY:  polyA_signal
;       LOCATION:  7630..7635
;       OTHER INFORMATION:  /standard_name= "Bovine beta casein
;       OTHER INFORMATION:  3' region, in exon 9"
US-08-723-306-5


  Query Match             2.3%;  Score 48.2;  DB 2;  Length 11093;
  Best Local Similarity   52.2%;  Pred. No. 0.00037;
  Matches  107;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy     697 aaaaatttccatttqttcqaaaatcagaqaqaacactggaatgttacgttcataacttac 756
           ||| |     || | |   |||||     ||||||    ||| | | ||| || | |||
Db    2278 AAAGAAGAGTATATTIATTAAAATTGCTCAGAACATCCAAATTTCAAGTTTATCATTTAT 2337

Qy     757 taaggattagtgtatatttttccaaccttqaaggcatgaaattctgqaqcttattattqaaa 816
             | || |  | |  |    ||  | ||  |||||||| | | ||   ||||  || |
Db    2338 CTIACAATATTTCAAAAATATTAAAATAGATACATGAAATACAGAAGTAAATTAAAGAGA 2397

Qy     817 aactactcaagttqqatqtqaatgcatcccqqcagggtattgaaqatqctqaaqaaacag 876
           || || |  | || |     ||    ||  ||  |  ||  |    |||  || |
Db    2398 AAGTATTTTATTTTGTAAAAAAAAATTCTAGGTTGGACAGGGAGTACCAGGAAACAAAAA 2457

Qy     877 caactcaaacttqtqgtqggacaqa 901
             | | |||  |||| |  ||||||
Db    2458 ACAATGAAAAATGTGATCTGACAGA 2482


RESULT   4
PCT-US96-10041-5
; Sequence 5, Application PC/TUS9610041
;  GENERAL INFORMATION:
;    APPLICANT:  White PhD, Kenneth
;    APPLICANT:  Morrey PhD, John
;    APPLICANT:  Reed, William
;    TITLE OF INVENTION:  Cassette for Expression of Lytic
;    TITLE OF INVENTION:  Peptides in Mammalian Transgenic Organisms
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Trask Britt and Rossa
;      STREET:  P.O. Box 2550
;      CITY:  Salt Lake City
;      STATE:  Utah
;      COUNTRY:  USA
;      ZIP:  84110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/10041
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sweigert PhD, Susan E
;      REGISTRATION NUMBER:  36,289
;      REFERENCE/DOCKET NUMBER:  2549
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  8015321922
;      TELEFAX:  8015319168
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11093 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Construct comprising
;      DESCRIPTION:  portions of Bos taurus beta casein gene and genes encoding
;      DESCRIPTION:  amphipathic peptide and green fluorescent protein"
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1801..1844
;      OTHER INFORMATION:  /product= "beta casein exon 1"
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  3780..3832
;      OTHER INFORMATION:  /product= "beta casein exon 2"
;    FEATURE:
;      NAME/KEY:  TATA_signal
;      LOCATION:  1766..1773
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  4567..4590
;      OTHER INFORMATION:  /product= "portion of beta casein
;      OTHER INFORMATION:  exon 3"
;    FEATURE:
;      NAME/KEY:  5'UTR
;      LOCATION:  127..1800
;      OTHER INFORMATION:  /function= "5' flanking regulatory
;      OTHER INFORMATION:  region of bovine beta casein gene"
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  4587..5310
;      OTHER INFORMATION:  /product= "Green fluorescent
;      OTHER INFORMATION:  protein"
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  5320..5449
;      OTHER INFORMATION:  /product= "Shiva-1 coding sequence"
;    FEATURE:
;      NAME/KEY:  polyA_signal
;      LOCATION:  7630..7635
;      OTHER INFORMATION:  /standard_name= "Bovine beta casein
;      OTHER INFORMATION:  3' region, in exon 9"
PCT-US96-10041-5


  Query Match             2.3%;  Score 48.2;  DB 5;  Length 11093;
  Best Local Similarity   52.2%;  Pred. No. 0.00037;
  Matches  107;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy     697 aaaaatttccatttqttcqaaaatcagaqaqaacactqqaatqttacqttcataacttac 756
           ||| |     || | |   |||||     ||||||    ||| | | ||| || | |||
Db    2278 AAAGAAGAGTATATTTATTAAAATTGCTCAGAACATCCAAATTTCAAGTTTATCATTTAT 2337

Qy     757 taaqqattagtqtatatttttccaaccttqaaqcatqaaattctqqaqcttattattqaaa 816
             | || |  | |  |    ||  | ||  |||||||| | | ||   ||||  || |
Db    2338 CTIACAATATTTCAAAAATATTAAAATAGATACATGAAATACAGAAGTAAATTAAAGAGA 2397

Qy     817 aactactcaagttqqatqtqaatgcatcccqqcaqqqtattgaaqatqctqaaqaaacag 876
           || || |  | || |     ||    ||  ||  |  ||  |    |||  || |
Db    2398 AAGTATTTTATTTTGTAAAAAAAAATTCTAGGTTGGACAGGGAGTACCAGGAAACAAAAA 2457
```

```
Qy        877 caactcaaacttgtggtgggacaga 901
                |  |  | | |   | | | |  |   | | | | | |
Db       2458 ACAATGAAAAATGTGATCTGACAGA 2482


RESULT 5
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pFZgpt-F1s
US-08-232-463-14

  Query Match            2.0%;  Score 40.4;  DB 1;  Length 7218;
  Best Local Similarity  6.8%;  Pred. No. 0.073;
  Matches   29;  Conservative  207;  Mismatches  188;  Indels    0;  Gaps    0;

Qy        584 tgatgatgaagatgataatcttcctgcaaatttttgacacatgtcacagagccttgcaaat 643
              | |  ||    ||| ||     |   |  | : :::      :: :  :  :   : ::::      : :::
Db       1458 TAAAGAGATAGAAGAATTTCGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1399

Qy        644 aatagcaagatatgtaccatcgacaccgtggtttctcatgccaatactggtggaaaaatt 703
              :: :: :::: :  :  :   :   :: :   : ::        : :   :: :   :: :::::::
Db       1398 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1339

Qy        704 tccatttgttcgaaaatcagagaacactggaatgttacgttcataacttactaaggat 763
                  :   :      :::::  :::::::: :   :::: :   : :    : ::     :  :::::
Db       1338 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1279

Qy        764 tagtgtatattttccaaccttgaggcatgaaattctggagcttattattgaaaaactact 823
              :: :  : :       ::      :::: : :::.      ::::     :   :   ::::::    :
Db       1278 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1219

Qy        824 caagttggatgtgaatgcatcccggcagggtattgaagatgctgaagaaacagcaactca 883
              :::   ::: : ::: :  :      :: ::::  :    ::::: :   :::::: :: ::    :
Db       1218 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1159

Qy        884 aacttgtggtgggacagattccacggaaggattgtttaatatggatgaagatgaagaaac 943
              .:      : ::  ::::  :::       :  :::::::    :      ::  :  :::  :::::  ::::::
Db       1158 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1099

Qy        944 tgaacatgaaacaaaggctggtcctgaacggctcgaccagatggtgcatcctgtagccga 1003
              ::: :  ::::  :::::   ::       ::: ::         | ||  |        |    |||| |||| |
Db       1098 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCTCCCTCGACCTGCAGCCAA 1039

Qy       1004 gcgc 1007
              || |
Db       1038 GCTC 1035


RESULT 6
US-08-0’8-682-3
; Seque’ce 3, Application US/08038682
; Patent No. 5549897
;  GENERAL INFORMATION:
;    APPLICANT:  BARENKAMP, STEPHEN J
;    APPLICANT:  ST. GEME III, JOSEPH W
;    TITLE OF INVENTION:  HIGH MOLECULAR WEIGHT SURFACE PROTEINS
;    TITLE OF INVENTION:  OF NON-TYPEABLE HAEMOPHILUS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/038,682
;      FILING DATE:  16-MAR-1993
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BERKSTRESSER, JERRY W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-293
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4937 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-0’8-682-3

  Query Match            1.9%;  Score 39.2;  DB 1;  Length 4937;
  Best Local Similarity  53.2%;  Pred. No. 0.14;
  Matches   83;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        803 gcttattattgaaaaactactcaagttggatgtgaatgcatcccggcagggtattgaaga 862
               | ||||||   | ||||        ||| |   |||     |||||||   |    |    |||   ||||
Db       4287 GCTTATTAACGCAAAAGATGCTAAGCTAAATGGTGATGCATCAGGTGATAGTACAGAAGT 4346

Qy        863 tgctgaagaaacagcaactcaaacttgtggtgggacagattccacggaaggattgtttaa 922
```

```
Db   4347 GAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTGACTGCGGCAACCTCAAGCAGTGTGAA 4406

Qy    923 tatgqatqaaqatqaaqaaactqaacatqaaacaaa 958
          |||   ||  ||| | | || | | |||    |||
Db   4407 TATCACTGGGGATTTAAACACAGTAAATGGGTTAAA 4442


RESULT   7
US-08-302-832-3
; Sequence 3, Application US/08302832
; Patent No. 5603938
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5603938-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/302,832
;      FILING DATE:  16-SEP-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US pct/us93/02166
;      FILING DATE:  16-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-404
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4937 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-302-832-3


  Query Match             1.9%;  Score 39.2;  DB 1;  Length 4937;
  Best Local Similarity   53.2%;  Pred. No. 0.14;
  Matches   83;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy     803 gcttattattgaaaaactactcaagttqqatqtqaatqcatcccqqcaqqqtattqaaqa 862
           | |||||||  | ||||      ||| |  |||   |||||||  |  |  |||  ||||
Db    4287 GGTTATTAACGCAAAAGATGCTAAGCTAAATGGTGATGCATCAGGTGATAGTACAGAAGT 4346

Qy     863 tqctqaaqaaacaqcaactcaaacttqtqqtqqqacaqattccacqqaaqqattqtttaa 922
             || ||  |  ||||      || || ||| ||| |   | ||   | |     | ||
Db    4347 GAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTGACTGCGGCAACCTCAAGCAGTGTGAA 4406

Qy     923 tatgqatqaaqatqaaqaaactqaacatqaaacaaa 958
           |||   ||  ||| | | || | | |||    |||
Db    4407 TATCACTGGGGATTTAAACACAGTAAATGGGTTAAA 4442


RESULT   8
US-08-530-198-3
; Sequence 3, Application US/08530198
; Patent No. 5869065
;  GENERAL INFORMATION:
;    APPLICANT:  BARENKAMP, STEPHEN J
;    APPLICANT:  ST. GEME III, JOSEPH W
;    TITLE OF INVENTION:  HIGH MOLECULAR WEIGHT SURFACE PROTEINS
;    TITLE OF INVENTION:  OF NON-TYPEABLE HAEMOPHILUS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/530,198
;      FILING DATE:  13-DEC-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BERKSTRESSER, JERRY W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  JWB-1186
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4937 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-530-198-3


  Query Match             1.9%;  Score 39.2;  DB 2;  Length 4937;
  Best Local Similarity   53.2%;  Pred. No. 0.14;
  Matches   83;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy     803 gcttattattgaaaaactactcaagttqqatqtqaatqcatcccqqcaqqqtattqaaqa 862
           | |||||||  | ||||      ||| |  |||   |||||||  |  |  |||  ||||
Db    4287 GGTTATTAACGCAAAAGATGCTAAGCTAAATGGTGATGCATCAGGTGATAGTACAGAAGT 4346

Qy     863 tqctqaaqaaacaqcaactcaaacttqtqqtqqqacaqattccacqqaaqqattqtttaa 922
             || ||  |  ||||      || || ||| ||| |   | ||   | |     | ||
Db    4347 GAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTGACTGCGGCAACCTCAAGCAGTGTGAA 4406

Qy     923 tatgqatqaaqatqaaqaaactqaacatqaaacaaa 958
           |||   ||  ||| | | || | | |||    |||
Db    4407 TATCACTGGGGATTTAAACACAGTAAATGGGTTAAA 4442


RESULT   9
US-08-469-880-3
; Sequence 3, Application US/08469880
; Patent No. 5876733
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J.
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5876733-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
```

```
;       STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;       STREET:  Bldg. 1
;       CITY:  Arlington
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22202-0286
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/469,880
;       FILING DATE:  06-JUN-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9205704.1
;       FILING DATE:  16-MAR-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US PCT/US93/02166
;       FILING DATE:  16-MAR-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/302,832
;       FILING DATE:  16-SEP-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Berkstresser, Jerry W
;       REGISTRATION NUMBER:  22,651
;       REFERENCE/DOCKET NUMBER:  1038-516 MIS:vq
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 415-0810
;       TELEFAX:  (703) 415-0813
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4937 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-469-880-3

  Query Match             1.9%;  Score 39.2;  DB 2;  Length 4937;
  Best Local Similarity   53.2%;  Pred. No. 0.14;
  Matches   83;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy     803 gcttattattgaaaaactactcaagttggatgtgaatgcatccccggcagggtattgaaga 862
                ||||||   | ||||           ||| |  |||    |||||||  |   |||  ||||
Db    4287 GGTTATTAACGCAAAAGATGCTAAGCTAAATGGTGATGCATCAGGTGATAGTACAGAAGT 4346

Qy     863 tgctgaagaaacagcaactcaaacttgtggtgggacagattccacggaaggattgtttaa 922
                    || ||  |  ||||        || || ||| ||| |    | ||    | |        | ||
Db    4347 GAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTGACTGCGGCAACCTCAAGCAGTGTGAA 4406

Qy     923 tatggatgaagatgaagaaactgaacatgaaacaaa 958
                |||    ||  |||  | | || | | |||      |||
Db    4407 TATCACTGGGGATTTAAACACAGTAAATGGGTTAAA 4442

RESULT 10
US-08-728-470-3
; Sequence 3, Application US/08728470
; Patent No. 5928651
;   GENERAL INFORMATION:
;     APPLICANT:  Barenkamp, Stephen J
;     TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;     TITLE OF INVENTION:  of No. 5928651 Typeable Haemophilus
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Shoemaker and Mattare, Ltd.
;       STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;       STREET:  Bldg. 1
;       CITY:  Arlington
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22202-0286
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/728,470
;       FILING DATE:
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/302,832
;       FILING DATE:  16-MAR-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US PCT/US93/02166
;       FILING DATE:  16-MAR-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9205704.1
;       FILING DATE:  16-MAR-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Berkstresser, Jerry W
;       REGISTRATION NUMBER:  22,651
;       REFERENCE/DOCKET NUMBER:  1038-633
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 415-0810
;       TELEFAX:  (703) 415-0813
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4937 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-728-470-3

  Query Match             1.9%;  Score 39.2;  DB 2;  Length 4937;
  Best Local Similarity   53.2%;  Pred. No. 0.14;
  Matches   83;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy     803 gcttattattgaaaaactactcaagttggatgtgaatgcatccccggcagggtattgaaga 862
                ||||||   | ||||           ||| |  |||    |||||||  |   |||  ||||
Db    4287 GGTTATTAACGCAAAAGATGCTAAGCTAAATGGTGATGCATCAGGTGATAGTACAGAAGT 4346

Qy     863 tgctgaagaaacagcaactcaaacttgtggtgggacagattccacggaaggattgtttaa 922
                    || ||  |  ||||        || || ||| ||| |    | ||    | |        | ||
Db    4347 GAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTGACTGCGGCAACCTCAAGCAGTGTGAA 4406

Qy     923 tatggatgaagatgaagaaactgaacatgaaacaaa 958
                |||    ||  |||  | | || | | |||      |||
Db    4407 TATCACTGGGGATTTAAACACAGTAAATGGGTTAAA 4442

RESULT 11
US-08-617-697-3
; Sequence 3, Application US/08617697
; Patent No. 5977336
;   GENERAL INFORMATION:
;     APPLICANT:  Barenkamp, Stephen J
;     TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;     TITLE OF INVENTION:  of No. 5977336-Typeable Haemophilus
;     NUMBER OF SEQUENCES:  11
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Shoemaker and Mattare, Ltd.
;       STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;       STREET:  Bldg. 1
;       CITY:  Arlington
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22202-0286
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
```

```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/617,697
;       FILING DATE:  01-APR-1996
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/302,832
;       FILING DATE:  05-OCT-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US PCT/US93/02166
;       FILING DATE:  16-MAR-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Berkstresser, Jerry W
;       REGISTRATION NUMBER:  22,651
;       REFERENCE/DOCKET NUMBER:  1038-557
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 415-0810
;       TELEFAX:  (703) 415-0813
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4937 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-617-697-3


  Query Match             1.9%;  Score 39.2;  DB 2;  Length 4937;
  Best Local Similarity   53.2%;  Pred. No. 0.14;
  Matches   83;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy     803 gcttattattgaaaaactactcaagttggatgtgaaatgcatcccggcaggqtattgaaga 862
           | |||||||  | ||||      ||| |  |||   ||||||||  |   |  |||  ||||
Db    4287 GGTTATTAACGCAAAAGATGCTAAGCTAAATGGTGATGCATCAGGTGATAGTACAGAAGT 4346

Qy     863 tgctgaaqaaacaqcaactcaaacttgtggtgqqacagattccacqqaaqqattqtttaa 922
             || ||  |  ||||      || || ||| ||| |   | ||  | |     | ||
Db    4347 GAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTGACTGCGGCAACCTCAAGCAGTGTGAA 4406

Qy     923 tatqqatqaaqatqaaqaaactqaacatqaaacaaa 958
           |||   ||  ||| | | || | | |||    |||
Db    4407 TATCACTGGGGATTTAAACACAGTAAATGGGTTAAA 4442


RESULT  12
US-08-719-641-3
; Sequence 3, Application US/08719641
; Patent No. 6218141
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 6218141-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/719,641
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-625
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4937 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-719-641-3


  Query Match             1.9%;  Score 39.2;  DB 4;  Length 4937;
  Best Local Similarity   53.2%;  Pred. No. 0.14;
  Matches   83;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy     803 gcttattattgaaaaactactcaagttggatgtgaaatgcatcccggcaggqtattgaaga 862
           ' |||||||  | ||||      ||| |  |||   ||||||||  |   |  |||  ||||
Db    4287 GGTTATTAACGCAAAAGATGCTAAGCTAAATGGTGATGCATCAGGTGATAGTACAGAAGT 4346

Qy     863 tgctgaaqaaacaqcaactcaaacttgtggtgqqacagattccacqqaaqqattqtttaa 922
             || ||  |  ||||      || || ||| ||| |   | ||  | |     | ||
Db    4347 GAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTGACTGCGGCAACCTCAAGCAGTGTGAA 4406

Qy     923 tatqqatqaaqatqaaqaaactqaacatqaaacaaa 958
           |||   ||  ||| | | || | | |||    |||
Db    4407 TATCACTGGGGATTTAAACACAGTAAATGGGTTAAA 4442


RESULT  13
US-08-231-193A-55
; Sequence 55, Application US/08231193A
; Patent No. 5849895
;  GENERAL INFORMATION:
;    APPLICANT:  Daggett, Lorrie P.
;    APPLICANT:  Ellis, Steven B.
;    APPLICANT:  Liaw, Chen W.
;    APPLICANT:  Lu, Chin-Chun
;    TITLE OF INVENTION:  HUMAN N METHYL-D-ASPARTATE RECEPTOR
;    TITLE OF INVENTION:  SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
;    NUMBER OF SEQUENCES:  63
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Martin, Haller & McClain
;      STREET:  1660 Union Street
;      CITY:  San Diego
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  92101-2926
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/231,193A
;      FILING DATE:  20-APR-1994
;      CLASSIFICATION:  536
;    PR     APPLICATION DATA:
;       PLICATION NUMBER:  US 08/052,459
```

```
;       FILING DATE:  20 APR 1993
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Seidman, Stephanie
;       REGISTRATION NUMBER:  33,779
;       REFERENCE/DOCKET NUMBER:  6362 9383
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  619 238 0999
;       TELEFAX:  619 238 0062
;   INFORMATION FOR SEQ ID NO:  55:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5548 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  both
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  210..4664
US 08 231 193A 55


  Query Match              1.7%;  Score 35.2;  DB 2;  Length 5548;
  Best Local Similarity   50.0%;  Pred. No. 2.5;
  Matches   88;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy    1762 qtqtqctqaaaqaqctcaaaqaaattcattqatcctatttatcaqqtatqqqaaaqacatqa 1821
           | ||| |||| || | | | | | | | |  | || | || |||      |||
Db    1333 GGGTGGGGAAGTGGAAAGACAAGTCCCTGCAGATGAAGTACTATGTGTGGCCCCGAATGT 1392

Qy    1822 qtqctqaaaqaqctacaqqaqttcaaaqaaacccatqaaaaaqqacataqtqqaaqatqaaq 1881
           || | ||      | || ||    | | | |  ||| | |  |  ||||  | | | |
Db    1393 GTCCAGAGACTGAAGAGCAGGAGGATGACCATCTGAGCATTGTGACCCTGGAGGAGGCAC 1452

Qy    1882 atqatqacttttctqaaaqqcqaaqtqcccccaqaatqataccqtqattqqqatcaca 1937
             || | || || || | |  | || | || ||  |||   ||  ||| ||||
Db    1453 CATTTGTCATTGTGGAAAGTGTGGACCCTCTGAGTGGAACCTGCATGAGGAACACA 1508


RESULT  14
US 08 486 273A 55
; Sequence 55, Application US/08486273A
; Patent No. 5985586
;  GENERAL INFORMATION:
;     APPLICANT:  Daggett, Lorrie P.
;     APPLICANT:  Ellis, Steven B.
;     APPLICANT:  Liaw, Chen W.
;     APPLICANT:  Lu, Chin Chun
;     TITLE OF INVENTION:  HUMAN N METHYL D ASPARTATE RECEPTRO SUBUNITS, DNA
;     TITLE OF INVENTION:  ENCODING SAME AND USES THEREFOR
;     NUMBER OF SEQUENCES:  64
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Brown, Martin, Haller & McClain
;       STREET:  1660 Union Street
;       CITY:  San Diego
;       STATE:  CA
;       COUNTRY:  U.S.A.
;       ZIP:  92101 2926
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/486,273A
;       FILING DATE:  06-JUN 1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/231,193
;       FILING DATE:  20 APR 1994
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Seidman, Stephanie
;       REGISTRATION NUMBER:  33,779
;       REFERENCE/DOCKET NUMBER:  6362 9383B
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  619-238-0999
;       TELEFAX:  619-238-0062
;   INFORMATION FOR SEQ ID NO:  55:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5548 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  both
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  210..4664
US 08 486 273A 55


  Query Match              1.7%;  Score 35.2;  DB 2;  Length 5548;
  Best Local Similarity   50.0%;  Pred. No. 2.5;
  Matches   88;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy    1762 qtqtqctqaaaqaqctcaaaqaaattcattqatcctatttatcaqqtatqqqaaaqacatqa 1821
           | ||| |||| || | | | | | | | |  | || | || |||      |||
Db    1333 GGGTGGGGAAGTGGAAAGACAAGTCCCTGCAGATGAAGTACTATGTGTGGCCCCGAATGT 1392

Qy    1822 qtqctqaaaqaqctacaqqaqttcaaaqaaacccatqaaaaaqqacataqtqqaaqatqaaq 1881
           || | ||      | || ||    | | | |  ||| | |  |  ||||  | | | |
Db    1393 GTCCAGAGACTGAAGAGCAGGAGGATGACCATCTGAGCATTGTGACCCTGGAGGAGGCAC 1452

Qy    1882 atqatqacttttctqaaaqqcqaaqtqcccccaqaatqataccqtqattqqqatcaca 1937
             || | || || || | |  | || | || ||  |||   ||  ||| ||||
Db    1453 CATTTGTCATTGTGGAAAGTGTGGACCCTCTGAGTGGAACCTGCATGAGGAACACA 1508


RESULT  15
US 08 940 986A 55
; Sequence 55, Application US/08940086A
; Patent No. 6111091
;  GENERAL INFORMATION:
;     APPLICANT:  Daggett, Lorrie P.
;     APPLICANT:  Ellis, Steven B.
;     APPLICANT:  Liaw, Chen W.
;     APPLICANT:  Lu, Chin-Chun
;     TITLE OF INVENTION:  HUMAN N METHYL D ASPARTATE RECEPTOR
;     TITLE OF INVENTION:  SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
;     NUMBER OF SEQUENCES:  64
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Heller Ehrman White & McAuliffe
;       STREET:  4250 Executive Square, 7th Floor
;       CITY:  La Jolla
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  92037
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/940,086A
;       FILING DATE:  29-SEPT 97
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/231,193
;       FILING DATE:  20-APR-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/052,449
;       FILING DATE:  20-APR-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Seidman, Stephanie
;       REGISTRATION NUMBER:  33,779
;       REFERENCE/DOCKET NUMBER:  24735 9383C
```

```
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 450-8400
;       TELEFAX:  (619) 450-8499
;   INFORMATION FOR SEQ ID NO:  55:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5538 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  both
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  210..4664
US-08-940-086A-55


  Query Match             1.7%;  Score 35.2;  DB 3;  Length 5538;
  Best Local Similarity   50.0%;  Pred. No. 2.5;
  Matches   88;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy    1762 gtgtgctgaagaggtcaaagaaattcattgatcctatttatcaggtatgggaagacatga 1821
           | |||  |||| ||  | | || | | |  |    |  ||  | || |||      |||
Db    1333 GGGTGGGGAAGTGGAAAGACAAGTCCCTGCAGATGAAGTACTATGTGTGGCCCCGAATGT 1392

Qy    1822 gtgctgaagagctacagggagttcaagaaacccatgaaaaaggacatagtggaagatgaag 1881
           || | ||      | || ||    |  | |   ||| |  |  |   |||| || | |
Db    1393 GTCCAGAGACTGAAGAGCAGGAGGATGACCATCTGAGCATTGTGACCCTGGAGGAGGCAC 1452

Qy    1882 atgatgactttctgaaaggcgaagtgccccagaatgataccgtgattgggatcaca 1937
             || | || || || | |  |  || | || ||  |||   ||  ||| ||||
Db    1453 CATTTGTCATTGTGGAAAGTGTGGACCCTCTGAGTGGAACCTGCATGAGGAACACA 1508


Search completed: July 30, 2002, 03:56:26
Job time: 4885 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 02:07:10 ; Search time 1723.16 Seconds
(without alignments)
16197.975 Million cell updates/sec

Title: US-09-932-678-1
Perfect score: 2068
Sequence: 1 acaqaqqctqtqqctqqaaq.....tqacatttqqqattcccat 2068

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: qb_est1:*
10: qb_est2:*
11: qb_htc:*
12: qb_qss:*
13: em_qss_hum:*
14: em_qss_inv:*
15: em_qss_pln:*
16: em_qss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	815.4	39.4	1800	10	Y16973	Y16973 RNY16973 Ra
2	675.4	32.7	788	10	BG502871	BG502871 602550430
3	666	32.2	733	10	BI561523	BI561523 603256222
4	653.6	31.6	996	10	BF203684	BF203684 601866331
5	646	31.2	757	10	BE734354	BE734354 601565596
6	637.4	30.8	1037	10	BG428305	BG428305 602498983
7	617.8	29.9	671	9	AV703279	AV703279 AV703279
8	614.6	29.7	736	10	BI464721	BI464721 603202341
9	582.2	28.2	990	10	BG167061	BG167061 602344740
10	553.4	26.8	828	10	BI761160	BI761160 603043657
11	550.4	26.6	742	10	BE870692	BE870692 601448528
12	548	26.5	550	9	AL597171	AL597171 DKFZp313K
13	547.2	26.5	727	10	BI549394	BI549394 603190049
14	542.4	26.2	563	9	AW408066	AW408066 UI-HF-BM0
15	536.8	26.0	564	9	AW958173	AW958173 EST370243
16	520.4	25.2	697	10	BG611364	BG611364 602612725
17	517	25.0	535	9	AA213789	AA213789 zr91a11.s
18	503.8	24.4	835	10	BI601238	BI601238 603245755
19	488.4	23.6	498	10	BF110176	BF110176 7n50d06.x
20	487.4	23.6	497	10	BF221544	BF221544 7o69d12.x
21	482.8	23.3	498	10	BG149493	BG149493 nad29e03.
22	476.6	23.0	503	9	AI799954	AI799954 we41d02.x
23	471.8	22.8	628	9	BB642011	BB642011 BB642011
24	470	22.7	490	10	BE502966	BE502966 hc81q07.x
25	465.8	22.5	946	10	BG535886	BG535886 602564040
26	463.2	22.4	632	10	BE614096	BE614096 601503834
27	461.4	22.3	1045	10	BG297178	BG297178 602395174
28	458	22.1	458	10	BE549643	BE549643 7b40q08.x
29	451.4	21.8	461	9	AI928274	AI928274 wo67c08.x
30	447.6	21.6	686	9	BB639379	BB639379 BB639379
31	440.4	21.3	956	10	BG295610	BG295610 602392992
32	438.4	21.2	440	9	AA811628	AA811628 ob74d03.s
33	428.8	20.7	807	10	BI913245	BI913245 603178980
34	428.4	20.7	807	10	BE271211	BE271211 600943821
35	421.8	20.4	547	9	AA530643	AA530643 vj44a04.r
36	420.4	20.3	422	9	BE093077	BE093077 RC5-BT074
37	419.2	20.3	811	10	BE560866	BE560866 601446151
38	409	19.8	421	9	AA191111	AA191111 zp86a03.r
39	406.4	19.7	479	9	AW239267	AW239267 xb38b04.y
40	399.2	19.3	529	9	AW968124	AW968124 EST380419
41	399.2	19.3	660	10	BI914307	BI914307 603180944
c 42	387	18.7	406	10	BI018122	BI018122 CM1-MT024
c 43	380	18.4	380	9	AA481295	AA481295 zv02e05.r
44	379.8	18.4	394	9	AW770928	AW770928 hl93e04.x
45	377.2	18.2	700	9	BB665490	BB665490 BB665490

ALIGNMENTS

RESULT 1
Y16973
LOCUS Y16973 1800 bp mRNA linear EST 30-APR 1999
DEFINITION RNY16973 Rat liver ESTs (E.Olivier) Rattus norvegicus cDNA clone IRL573, mRNA sequence.
ACCESSION Y16973
VERSION Y16973.2 GI:4727035
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 1800)
AUTHORS Olivier,E., Soury,E., Risler,J.L., Smih,F., Schneider,K., Lochner,K., Jouzeau,J.Y., Fey,G. and Salier,J.P.
TITLE A novel set of hepatic mRNAs preferentially expressed during an acute inflammation in rat represents mostly intracellular proteins
JOURNAL Genomics 57 (3), 352-364 (1999)
MEDLINE 99263497
COMMENT On Apr 2, 1998 this sequence version replaced gi:3006069.
Contact: E. Olivier
U78 INSERM
543 chemin de la Breteque, 76233 Bois-Guillaume, France
POLYA-NO.
FEATURES Location/Qualifiers
source 1. .1800
/organism="Rattus norvegicus"
/strain "Sprague-Dawley"
/db_xref="taxon:10116"
/clone="IRL573"
/clone_lib="Rat liver ESTs (E.Olivier)"
/tissue_type="liver"
/dev_stage="adult"
/note="Organ: liver; Vector: lambda ZAPII; Library construction: Ripperger,J.A. et al. J. Biol. Chem. (1995) 270(50):29998-30006"
BASE COUNT 473 a 393 c 433 g 501 t
ORIGIN

```
  Query Match             39.4%;  Score 815.4;  DB 10;  Length 1800;
  Best Local Similarity   75.7%;  Pred. No. 2e-221;
  Matches 1411;  Conservative    0;  Mismatches  366;  Indels   86;  Gaps   29;

Qy        135 aagctgggcgcgtcgaggactgggatttcaaatatgcgtgcattagagaatgactttttc 194
              | |||| | |||||| ||||||| |||||| |||| ||||||||||| ||| || |
Db         23 ACGCTGAGTGCGTCGCGGACTGGGCTTTCAAACATGCTTGCATTAGAGAGTGATTTCTCA 82

Qy        195 aattctccccaagaaaaactgttcggtttggtggaactgtgacagaagtcttgctgaag 254
                 || |     ||               |      | |   | ||||||||||| ||||| ||||
Db         83 GTCTCCCAAGAAAC------TGTCGGTTGGTGGACTGTGACAGACGTCTTACTGAGT 133

Qy        255 tacaaaaagggtgaaacaaatgactttgagttgttgaagaaccagctgttagatccagac 314
                 | ||| ||    ||||||||||| |||||||||||||| |||||||   ||||| ||
Db        134    ACAAAAGGGTGAACAAATGACTTAGAGTTGTTGAAGAA-CAGCTGTCTGATCCTGAT 189

Qy        315 ataaaggatgaccagatcatcaactggctgctagaattccgttcttctatcatgtacttg 374
              ||| ||||||| ||||||| |||||||||| ||||| ||||| ||| ||||||||||
Db        190 ATAGAGGATGATCAGATCATTAACTGGCTGCTGGAATTTCGTTCCTCTGTCATGTACTTG 249

Qy        375 acaaaagactttgagcaacttatcagtattatattaagattgccttggttgaatagaagt 434
              |||||||||||||||||||||| |   || ||||| |||||||| ||||||||||||||
Db        250 ACAAAAGACTTTGAGCAACTTATAAACATCATATTGAGATTGCCCTGGTTGAATAGAAGT 309

Qy        435 caaacagtagtggaagagtatttggctttcttggtaatcttgtatcagcacagactgtt 494
              || || || ||||| |||||||||||||||||||||||||||||| ||||||||||||||
Db        310 CAGACGGTGGTGGAGGAGTATTTGGCTTTTCTTGGTAATCTTGTGTCAGCACAGACTGTC 369

Qy        495 ttcctcagaccgtgtctcagcatgattgcttcccatttgtgcctccccgagtgatcatt 554
              ||||| ||||| |||||||||||||||||||| ||||||||| |||||||||| ||   |
Db        370 TTCCTTAGACCATGTCTCAGCATGATTGCTTCTCATTTTGTACCTCCCCGAGTTATTGTC 429

Qy        555 aaggaaggcgatgtagatgtttcagattctgatgatgaagatgataatcttcctgcaaat 614
              |||||||| |    | |||| |||||| ||||| ||||||||||| ||||||||||||||
Db        430 AAGGAAGGTGGCATTGATG-TTCAGAGTCTGAGGATGAAGATGACAATCTTCCTGCAAAT 488

Qy        615 tttgacacatgtcacagagccttgcaaataatagcaagatatgtaccatcgacaccgtgg 674
              |||||||||||||||| |||||||||||||||| |||||||||| || |||||||| |||
Db        489 TTTGACACATGTCACAAAGCCTTGCAAATAATAACAAGATATGTCCCTTCGACACCATGG 548

Qy        675 tttctcatgccaatactggtggaaaaatttccatttgttcgaaaatcagagagaacactg 734
              ||||| |||||||||||||| ||||||||||||||||    | ||||||||||||||||
Db        549 TTTCTAATGCCAATACTGGTAGAAAAATTTCCATTTGTGAGGAAATCAGAGAGAACACTG 608

Qy        735 gaatgttacgttcataacttactaaggattagtgtatatttccaaccttgaggcatgaa 794
              ||||| || ||||||||| || ||||||| ||| |||||||| ||||| ||||||| ||||
Db        609 GAATGCTATGTTCATAACCTATTAAGGATAAGTTTATATTTCCCAACGTTGAGGCGTGAA 668

Qy        795 attctggagcttattattgaaaaactactcaagttggatgtgaatgcatccgggcagggt 854
              ||||| ||||||||||||||||   | ||||||||| ||||||| ||   || ||||||| |
Db        669 ATTCTCGAGCTTATTATTGAAAGTC-ACTCAAGTTAGATGTGAGTGTGTCGCGGCAGGAT 727

Qy        855 atttgaaagatgctgaagaaacagcaactcaaacttgtggtgggacagattccacggaagga 914
              ||||| |||||||||| |||||| |||| |||| ||||||||||||   || | ||||||
Db        728 ATTGACGATGCTGAAGAGACAGCAGCTCAGACTTATGGTGGGACAGACGCCGGCAGAAGGA 787

Qy        915 ttgtttaatatggatgaagatgaagaaactgaacatgaaacaaaggctggtcctgaacgg 974
                || |||||||||||||       || |||||| | ||| |   || ||||  |   || ||
Db        788 CTG-TTAATATGGATGA---GATGAGGAACTGACCCTGAGAAGAAAGCTGAGCAGGAGCGC 844

Qy        975 ctcgaccagatggtgcatcctgtagccgagcgcctggacatcctgatgtctttggttttg 1034
              || | || ||||| |||||      || ||||||||||||||||||| |||| ||| |  |
Db        845 CT-GGCCGGATGGCTCATCCCACTGCAGAGCGCCTGGACATCCTGCTGTCCTTGCTGCTC 903

Qy       1035 tcctacatgaaggatgtctgctatgtagatggtaaggttgataacggcaaaacaaaggat 1094
              ||||||||  | ||||||||| ||||   |    ||||    ||||||||     ||||||||
Db        904 TCCTACATTGACGATGTCTGCCATGTGCA-GTAAAATTGATAACAATAAAACAAAG--A 959

Qy       1095 ctatatcgcgacctgataaacatctttgacaaactcctgttgcccacccatgcctcctgc 1154
               ||||||| || ||||||    |||||     | |      ||||||||| ||||||| ||||
Db        960 TTATATCGTGATCTGATATCTATCTT---GACAACTGTGTTGCCCACACATGCCT-CTGC 1015

Qy       1155 catgtacagttttcatgttttacctctgtagtttcaaattgggattcgcagaggcattt 1214
              |||||||||| |||||||||| |||          ||||||||| || ||||| ||||||
Db       1016 CATGTACAGTTCTTCATGTTTTTCCTTGCAGC---TCAAATTGGGGTTTGCAGA-GCATTT 1072

Qy       1215 ttggaacatctctggaaaaaattgcaggacccaagtaatcctgccatcatcaggcagct 1274
              ||||||||||||| |||       | |         ||| ||| |||| |||||||||      |||
Db       1073 TTGGAACATCTCTGAAAAGTGCAGGA----TCAAATAACCCTGTCATCATCAG---CAGCT 1126

Qy       1275 gctggaaattatattggaagcttttggcaagagctaaatttattcctcttattactgta 1334
              |||| ||||          |||||||||||  |     ||   ||| |||||||| |||||
Db       1127 GCTGCAAATACATC---AGCTTTTTGGCAGAACTAAA-TTAATCCTCTTATCACTGTG 1180

Qy       1335 aaatcatgcctagatctttggttaactggctgcacatataccttaataaccaggattcg 1394
               || |||||||| |||| ||||| |||||||||||| | |||||||||| ||||| |   |
Db       1181 AAGTCATGCCTGGATC-TTTGGGTAACTGGCTGCAC-TGTACCTTAATAGCCAGGTTCGG 1238

Qy       1395 ggaacaaaggcattctgcgatgttgctctccatggaccatttactcagcctgccaagct 1454
                |        |  ||  |   |       ||||||||||||  |       ||||||||||
Db       1239 ACA--------AGGCTTTGTGATGTGCACTCATGGACCATTTATC---AGCTGCCAAGCT 1287

Qy       1455 gtgttctacaccttgttttagacacaagcagctttgagcgggaaacctgaaagaaggt 1514
              ||||       |  |  ||| |||||||||||       |         |         || | ||
Db       1288 GTGT------CTACACTGTGTTTAGACACAGCACTTT-------GAGTGCACCTGAGGAAGT 1336

Qy       1515 ttgcagtatcttcagagtctgaatttgagcggatagtgatgagccagctaaatcccctg 1574
                | | |||||| |||||||| |||||||| || || ||| |||||||||| || || |||
Db       1337 CTACGGTATCTCCAGAGTCTAAATTTTGAACGCATTGTGCTGAGCCAGCTGAACCCACTG 1396

Qy       1575 aagatttgcctgccctcagtggttaactttttgctgcaatcacaaataagtaccagctc 1634
              ||||| ||||||||       |||   |   |||   |     | ||||  |  | |    ||  | |||
Db       1397 AAGATCTGCCTGCCACGTGTGTATTCCTTGCTGCATACAATAATACAGCTGTGTCTGCTA 1456

Qy       1635 gtctttctgctacaccatcattgagaggaacaatcgccagatgctgccagtcattaggagt 1694
                | ||      | |                | ||||||||||||||||| || || ||     ||||
Db       1457 CACATCATGGAGA------------GGACAATCGCCAGATGCTCCCTGTTAT--CGAGT 1501

Qy       1695 accgctggaggagactcagtgcagatctgcacaaacccgctggacaccttcttccccttt 1754
              || ||||| ||||||| ||||||| |||||| |||||   ||||||| ||||| ||||||
Db       1502 ACAGCTGGTGGAGACTCTGTGCAGACCTGCACCAACCCATTGGACACTTTCTTCCCTTT 1561

Qy       1755 gatccctgtgtgctgaagaggtcaaagaaattcattgatcctatttatcaggtatgggaa 1814
              || || ||| |||| ||||||||||||| ||||| |||||| | |||||||   ||||||
Db       1562 GACCCTTGTATGCTTAAGAGGTCAAAGAAGTTCATCGATCCTGTCTATCAGGCTGGGAA 1621

Qy       1815 gacatgagtgctgaagagctacaggagttcaagaaacccatgaaaaaggacatagtggaa 1874
              |||||||||| ||||||||| |||||| |   || ||||||    | ||||||   ||||||
Db       1622 CACATGAGTGATGAAGAGCCTGCAGGAGCT-TAGGAACCCACTAGAAAGGAGCTAGTGGAG 1680

Qy       1875 gatgaagatgatgactttctgaaaggcgaagtgccccagaatgataccgtgattgggatc 1934
              |||||||| ||||||||| |||| |       |     |||||| ||| || |||   |||   |
Db       1681 GATGAAGACGATGACTTTTTGAAGGCGAGTG---CCCAGAGTGACACAGTGGCTGGCCTT 1737

Qy       1935 acaccaagctcctttgacacgcatttccgaaagtccttcaagtagtgtgggctccccaccc 1994
              || || | ||||||||||| || || || ||||||||  |||||||||||||||||| ||
Db       1738 ACCCGAACTCCTTTGACACCCACTTTCG-AGTCCTTCCAGTAGTGTGGGCTCCCCTCCT 1796

Qy       1995 gtg 1997
              |||
Db       1797 GTG 1799


RESULT 2
BG502871
LOCUS       BG502871                788 bp    mRNA    linear   EST 27 MAR 2001
DEFINITION  602550430F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4658071 5',
            mRNA sequence.
ACCESSION   BG502871
VERSION     BG502871.1  GI:13464388
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 788)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1451  row: 1  column: 08
            High quality sequence stop: 699.
FEATURES             Location/Qualifiers
     source          1. .788
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4658071"
                     /clone_lib="NIH_MGC_61"
                     /tissue_type="embryonal carcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT       204 a    178 c    176 g    228 t      2 others
ORIGIN
```

```
  Query Match           32.7%;  Score 675.4;  DB 10;  Length 788;
  Best Local Similarity   96.1%;  Pred. No. 1.4e-181;
  Matches  756;  Conservative    0;  Mismatches   23;  Indels    8;  Gaps    6;

Qy     972 cggctcgaccagatggtgcatcctgtagccgagcgcctggacatcctgatgtctttggtt 1031
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 CGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCGCCTGGACATCCTGATGTCTTTGGTT 61

Qy    1032 ttgtcctacatgaaagatgtctgctatgtagatggtaaggttgataacggcaaaacaaag 1091
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 TTGTCCTACATGAAGGATGTCTGCTATGTAGATGGTAAGGTTGATAACGGCAAAACAAAG 121

Qy    1092 gatctatatcgcgaccctgataaacatctttgacaaactcctgttgcccacccatgcctcc 1151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 GATCTATATCGCGACCTGATAAACATCTTTGACAAACTCCTGTTGCCCACCCATGCCTCC 181

Qy    1152 tgccatgtacagtttttcatgttttacctctgtagtttcaaattgggattcgcagaggca 1211
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 TGCCATGTACAGTTTTTCATGTTTTACCTCTGTAGTTTCAAATTGGGATTCGCAGAGGCA 241

Qy    1212 tttttggaacatctctggaaaaaattgcaggacccaagtaatcctgccatcatcaggcag 1271
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 TTTTTGGAACATCTCTGGAAAAAATTGCAGGACCCAAGTAATCCTGCCATCATCAGGCAG 301

Qy    1272 gctgctggaaattatattggaagctttttggcaagagctaaatttattcctcttattact 1331
           ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     302 GCTGCTGGAAATTATATTGGAAGCTTTNTGGCAAGAGCTAAATTTATTCCTCTTATTACT 361

Qy    1332 gtaaaatcatgcctagatcttttggttaactggctgcacatatacctaataaccaggat 1391
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 GTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCAGGAT 421

Qy    1392 tcgggaacaaaggcattctgcgatgttgctctccatggaccatttacttcagcctgccaa 1451
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 TCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTACTTCAGCCTGCCAA 481

Qy    1452 gctgtgttctacacctttgttttttagacacaagcagctttttgagcggaaaacctgaaagaa 1511
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 GCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGAGCGGAAAACCTGAAAGAA 541

Qy    1512 ggtttgcaagtatcttcaagatctgaatttttgaagcggatagtgatgagccagctaaaigcc 1571
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     542 GGTTNGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAATGCC 601

Qy    1572 ctgaagatttgcctgccctcagtggttaacttttttgctgcaatcacaaa taagtacca 1630
           |||||||||||| ||||||||||||||||||||||||||||||||||||| |||||||||
Db     602 CTGAAGATTTG-CTGCCCTCAGTGGTTAACTTTTTTGCTGCAATCACAAATTAAGTACCA 660

Qy    1631 gctcgtcttctgttacaccatcatt-gaggaggaacaatcgccaga--tgctgccagtcat 1687
           |||||||||||| |||||||||||| |||||||||||| |||   | ||:   |||
Db     661 GCTCGTCTTCTGTTACACCATCATTGGAGAGGAACAATCGGCAGAATGGTTGCAGTCCAT 720

Qy    1688 taggagtaccgctggaggag-actcagtgccaga--tctgcacaaaccccgctggacacctt 1744
           | |||||||||||||||||| |||||||| ||     |   ||||||| |||| ||| |
Db     721 TGGGAGTACCGCTGGAGGAGAACTCAGTGCCGAATCTGGGCAAAACCCGGTGGAAACCCT 780

Qy    1745 cttcccc 1751
             || |||
Db     781 TTTCCC 787
```

```
RESULT   3
BI561523
LOCUS       BI561523                 733 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603256222F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5298581 5',
            mRNA sequence.
ACCESSION   BI561523
VERSION     BI561523.1  GI:15448837
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 733)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11755  row: 1  column: 06
            High quality sequence stop: 717.
FEATURES             Location/Qualifiers
     source          1. .733
                     /organism "Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5298581"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgac
                     ); oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3';
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT       205 a    142 c    171 g    214 t      1 others
ORIGIN
```

```
  Query Match             32.2%;  Score 666;  DB 10;  Length 733;
  Best Local Similarity   97.5%;  Pred. No. 6.4e-179;
  Matches  697;  Conservative    0;  Mismatches   16;  Indels    2;  Gaps    2;

Qy         53 gcgttagttcggcccaatggcggcaccgctgcttcacacgcgtttgccgggagatgcggc 112
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          7 GGGTTAGTTCGGCCCAATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGATGCGGC 66

Qy        113 cgcttcgtcctctgcagttaagaagctgggcgcgtcgaggactgggatttcaaatatgcg 172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 CGCTTCGTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATATGCG 126

Qy        173 tgcattagagaatgactttttcaattctcccccaagaaaaactgttcggtttggtggaac 232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 TGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTTCGGTTTGGTGGAAC 186

Qy        233 tgtgacagaagtcttgctgaagtacaaaaagggtgaaacaaatgactttgagttgttgaa 292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 TGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAA 246

Qy        293 gaaccagctgttagatccagacataaaggatgaccagatcatcaactggctgctagaatt 352
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db        247 GAACCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGTTGCTAGAATT 306

Qy        353 ccgttcttctatcatgtacttgacaaaagactttgagcaacttatcagtattatattaag 412
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 CCGTTCTTCTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAG 366

Qy        413 attgccttggttgaatagaagtcaaacagtagtggaagagtatttggcttttcttggtaa 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 ATTGCCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTGGTAA 426

Qy        473 tcttgtatcagcacagactgttttcctcagaccgtgtctcagcatgattgcttcccattt 532
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 TCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTT 486

Qy        533 tgtgcctccccgagtgatcattaaggaaggcgatgtagatgtttcagattctgatgatga 592
              ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        487 TGTGCCTTCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGATGATGA 546

Qy        593 agatgataatcttcctgcaaattttgacacatgtcacagagccttgcaaataatagcaag 652
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        547 AGATGATAATCTTCCTGCAAA-TTTGACACATGTCACAGAGCCTTGCAAATAATAGCAAG 605

Qy        653 atatgtaccatcgacaccgt-ggtttctcatgccaatactggtggaaaaatttccatttg 711
              |||||||||||||||||||| |||||||||||||||||||  |   | ||||  ||  | ||
Db        606 ATATGTACCATCGACACCGTGGGTTTCTCATGCCAATACTNGGTGGAAAAATTTCCATTT 665

Qy        712 ttcgaaaatcagagagaacactggaatgttacgttcataacttactaaggattag 766
              |    ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 GTTCGAAATCAGAGAGAACACTGGAATGTTACGTTCATAACTTACTAAGGATTAG 720


RESULT      4
BF203684
LOCUS       BF203684                     996 bp    mRNA    linear   EST 06-NOV-2000
DEFINITION  601866331F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4099398 5',
            mRNA sequence.
ACCESSION   BF203684
VERSION     BF203684.1  GI:11097270
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 996)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs r mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM967  row: n  column: 07
            High quality sequence stop: 684.
FEATURES             Location/Qualifiers
     source          1..996
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:4099398"
                     /clone_lib-"NIH_MGC_17"
                     /tissue_type "rhabdomyosarcoma"
                     /lab_host "DH10B (phage resistant)"
                     /note "organ: muscle; Vector: pOTB7; Site 1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size selected >500bp
                     for average insert size 1.8kb. Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      338 a    211 c    231 g    216 t
ORIGIN


  Query Match             31.6%;  Score 653.6;  DB 10;  Length 996;
  Best Local Similarity   97.0%;  Pred. No. 2.6e-175;
  Matches  687;  Conservative    0;  Mismatches   19;  Indels    2;  Gaps    2;

Qy       1202 cgcagaggcatttttggaacatctctggaaaaaattgcaggacccaagtaatcctgccat 1261
              |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db          2 CGCAGAGGCATTTTTGGAACATCTCTGGAAAAAA-TGCAGGACCCAAGTAATCCTGCCAT 60

Qy       1262 catcaggcaggctgctggaaattatattggaagctttttggcaagagctaaatttattcc 1321
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db         61 CATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTCGGCAAGAGCTAAATTTATTCC 120

Qy       1322 tcttattactgtaaaatcatgcctagatcttttggttaactggctgcacatataccttaa 1381
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        121 TCTTATTACTGTAAAATCATGCCTAGATCTTGTGGTTAACTGGCTGCACATATACCTTAA 180

Qy       1382 taaccaggattcgggaacaaaggcattctgcgatgttgctctccatggaccatttactc 1441
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTACTC 240

Qy       1442 agcctgccaagctgtgttctacacctttgtttttagacacaagcagctttgagcggaaa 1501
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTGAGCGGAAA 300

Qy       1502 cctgaaagaaggtttgcagtatcttcagagtctgaatttttgagcggatagtgatgagcca 1561
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCA 360

Qy       1562 gctaaatcccctgaagatttgcctgccctcagtggttaactttttttgctgcaatcacaaa 1621
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGCAATCACAAA 420

Qy       1622 taagtaccagctcgtcttctgctacaccatcattgagaggaacaatcgccagatgctgcc 1681
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAATCGCCAGATGCTGCC 480

Qy       1682 agtcattaggagtaccgctggaggagactcagtgcagatctgcacaaacccgctggacac 1741
              ||||||||||||||||||||||||||||||||||||||||||    || ||     |   ||
Db        481 AGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGACAAACCCGCTGGACAAC 540

Qy       1742 ctttcttcccctttgatccctgtgtgctgaagaggtcaaagaaattcattgatcctattta 1801
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTTCTTCCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTTA 600

Qy       1802 tcaggtatgggaagacatgagtgctgaagagctacaggagttcaagaaacccatgaaaaa 1861
```

```
          |||||| |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    601 TCAGGTGTGGGAAGACATGAGTGCTGAAGAGCTACAGGAGTTCAAGAAAACCATGAAAAA 660

Qy   1862 ggacatagtggaaagatgaaagatgatgactttctgaaaaggcgaagtgcc 1909
          |||||||||||||||||||||||||| | |||||||||| ||| |||
Db    661 GGACATAGTGGAAGATGAAGATGATG-CATTCTGAAAGGAGAATGGCC 707


RESULT   5
BE734354
LOCUS       BE734354                757 bp    mRNA    linear   EST 15-SEP-2000
DEFINITION  601565596F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3840623 5',
            mRNA sequence.
ACCESSION   BE734354
VERSION     BE734354.1  GI:10148346
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 757)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM530  row: o  column: 24
            High quality sequence stop: 693.
FEATURES             Location/Qualifiers
     source          1. .757
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3840623"
                     /clone_lib="NIH_MGC_21"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      233 a    152 c    166 g    206 t
ORIGIN


  Query Match           31.2%;  Score 646;  DB 10;  Length 757;
  Best Local Similarity   95.9%;  Pred. No. 3.4e-173;
  Matches  727;  Conservative    0;  Mismatches   25;  Indels    6;  Gaps    6;

Qy    607 ctgcaaatttgacacatgtcacagagccttgcaaataatagcaagatatgtaccatcga 666
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db      1 CTGCAAATTTGACACATGTCACAGAGCCTTGCAACTAATAGCAAGATATGTACCATCGA 60

Qy    667 caccgtg-gtttctcatgccaatactggtggaaaaatttccatttgttcgaaaatcagag 725
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CACCGTGTATTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTCGAAAATCAGAG 120

Qy    726 agaacactgga-atgttacgttcataacttactaaggattagtgtatatttttccaacctt 784
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGAACACTGGACATGTTACGTTCATAACTTACTAAGGATTAGTGTATATTTTTCCAACCTT 180

Qy    785 gaggcatgaaattctggagcttattattgaaaaactactcaagttggatgtgaatgcatc 844
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GAGGCATGAAATTCTGGAGCTTATTATTGAAAAACTACTCAAGTTGGATGTGAATGCATC 240

Qy    845 cggqcaggg-tattgaaagatgctgaaggaacaqcaactcaaaacttgtggtgggacggatt 903
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCGGCAGGGCTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGGTGGGACAGATT 300

Qy    904 ccacggaaggattgtttaatatggatgaagatgaagaaactgaacatgaaaacaaa ggct 962
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    301 CCACGGAAGGATTGTTTAATATGGATGAAGATGAAGAAACTGAACATGAAACAAACGGCT 360

Qy    963 ggtcctgaacggctcgaccagatggtgcatcctgtagccgaggcgcctggacatcctgatg 1022
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGTCCTGAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGGCGCCTGGACATCCTGATG 420

Qy   1023 tctttggtttgtcctacatgaaggatgtctgctatgtagatggtaaggttgataacggc 1082
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCTTTGGTTTGTCCTACATGAAGGATGTCTGCTATGTAGATGGTAAGGTTGATAACGGC 480

Qy   1083 aaaacaaaggatctatatcgcgacctgataaacatctttgacaaactcctgtagcccaac 1142
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    481 AAAACAAAGGATCTATATCGCGACCTGATAAACATCTTTGACAAACTCCTGTAGCCCAAC 540

Qy   1143 catgcctcctgccatgtacagttttttcatgttttacctctgtagtttcaaatgggattc 1202
          |||||||||||||||||||||||| || ||||| ||||||||||||||||||||| ||||
Db    541 CATGCCTCCTGCCATGTACAGTATTACATGTATTACCTCTGTAGTTTCAAATGGGATTC 600

Qy   1203 gcagaggcatttttgggaacatctctgaaaaaaattgcaggacccaagtaatcctgccatc 1262
          |||||||||| || ||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GCAGAGGCATATTAGGAACATCTCTGGAAAAAATTGCAGGACCCAAGTAATCCTGCCATC 660

Qy   1263 atcaggcaggctgctgtggaaattatattgggaagctttttggcaaagagctaaatttattcct 1322
          ||||||||||||||||||||||||| |||||| || ||| |||||| |||| ||||||
Db    661 ATCAGGCAGGCTGCTGGAAATTATATGGGAAGC-TTATGGGAAGAGCTAATTAATTCCT 719

Qy   1323 cttattactgtaaaatcatgcctagatcttttgqttaa 1360
          | ||||||||    |||||||||||   || || |||
Db    720 C-TATTACTGGTCCATCATGCCTAGAATCTTAGGATAA 756


RESULT   6
BG428305
LOCUS       BG428305                1037 bp    mRNA    linear   EST 14-MAR-2001
DEFINITION  602498983F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4612605 5',
            mRNA sequence.
ACCESSION   BG428305
VERSION     BG428305.1  GI:13334811
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1037)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1361  row: e  column: 22
            High quality sequence stop: 637.
FEATURES             Location/Qualifiers
     source          1. .1037
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4612605"
                     /clone_lib="NIH_MGC_75"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
```

```
                 StiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                 3' adaptors were used in cloning as follows: 5' adaptor
                 sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                 5'-ATTCTAGAGGCCGAGGCGGCCGACATG dT(30)BN-3' (where B = A,
                 C, or G and N = A, C, G, or T). Average insert size 1.65
                 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
                 by PCR. This library was enriched for full-length clones
                 and was constructed by Clontech Laboratories (Palo Alto,
                 CA). Note: this is a NIH MGC Library."
BASE COUNT      292 a    218 c    263 g    264 t
ORIGIN

  Query Match             30.8%;  Score 637.4;  DB 10;  Length 1037;
  Best Local Similarity   88.7%;  Pred. No. 1.2e-170;
  Matches  797;  Conservative    0;  Mismatches   81;  Indels   21;  Gaps    9;

Qy         49 ggtcgcgttagttcggcccaatggcggcaccgctgcttcacacgcgtttgccgggagatg 108
Db          2 GGTCGCGTTAGTTCGGCCCAATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGATG 61

Qy        109 cggccgcttcgtcctctgcagttaagaagctgggcgcgtcgaggactgggatttcaaata 168
Db         62 CGGCCGCTTCGTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATA 121

Qy        169 tgcgtgcattagagaatgactttttcaattctcccccaagaaaaactgttcggtttggtg 228
Db        122 TGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTTCGGTTTGGTG 181

Qy        229 gaactgtgacagaagtcttgctgaagtacaaaaagggtgaaacaaatgactttgagttgt 288
Db        182 GAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTGT 241

Qy        289 tgaagaaccagctgttagatccagacataaaggatgaccagatcatcaactggctgctag 348
Db        242 TGAAGAACCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAG 301

Qy        349 aattccgttcttctatcatgtacttgacaaaagactttgagcaacttatcagtattatat 408
Db        302 AATTCCGTTCTTCTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATAT 361

Qy        409 taagattgccttggttgaatagaagtcaaacagtagtggaagagtatttggctttcttg 468
Db        362 TAAGATTGCCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTG 421

Qy        469 gtaatcttgtatcagcacagactgtttttcctcagaccgtgtctcagcatgattgcttccc 528
Db        422 GTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTCCC 481

Qy        529 atttgtgcctcccg-agtgatcattaaggaaggcgatgtagatgtttcagattctgat 587
Db        482 ATTTGTGCCTCCCCGAAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGAT 541

Qy        588 gatgaagatgataatcttcctgcaaatttttgacacatgtcacagagccttgcaaataata 647
Db        542 GATGAAGATGATAATCTTCCTGCAAA-TTTGACACATGTCACAGAGCCTTGCAAATAATA 600

Qy        648 gcaagatatgtaccatcgacaccg-tggtttctcatgccaatactgg--tggaaaaattt 704
Db        601 GCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGGCAACAATTC 660

Qy        705 ccatttgttcgaaaa      tcagagagaacactggaatgtt--acgttcat--aacttac 756
Db        661 CCATATGGTCAGAAAAACTCCAGAGAGAACACTGGAACTGTCACCGTTCATGAACTTACT 720

Qy        757 taaggattagtgtatattttccaaccttg-aggcatgaaattctggagcttattattga 814
Db        721 TCAGGATCACGGGATATGTCCCAAGCTTGGAGGCATGCAAATTCTGGAGCTTATTATTGA 780

Qy        815 aaaactactcaagttggatgtgaatgcatccccggcagggtattgaagatg-----ctga 868
Db        781 AAACTACTCAAGGTGGACTGTGAATGGCTCCCGGGGCGGGTATTTGAAGACTGCTGAGG 840

Qy        869 ggaaacagcaactcaaacttgtggtgggacagattccacggaaggattgtttaatatgg 927
Db        841 AAAGCACGAAGTACAAACTGGTGCGGGAACAAATCCCGGCAAGGGATTGTTCAGTTTGG 899


RESULT 7
AV703279
LOCUS       AV703279                 671 bp    mRNA    linear   EST 09-OCT-2000
DEFINITION  AV703279 ADB Homo sapiens cDNA clone ADBCCD12 5', mRNA sequence.
ACCESSION   AV703279
VERSION     AV703279.1  GI:10720608
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 671)
  AUTHORS   Peng,Y., Song,H., Huang,Q., Huang,C., Gu,Y., Yang,Y., Gao,G., Xiao
            ,H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z.,
            Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
            ,G., Hu,R., Chen,J., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA ADB clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1..671
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="ADBCCD12"
                     /clone_lib="ADB"
                     /tissue_type="Adrenal gland"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      189 a    131 c    157 g    193 t      1 others
ORIGIN

  Query Match             29.9%;  Score 617.8;  DB 9;  Length 671;
  Best Local Similarity   97.0%;  Pred. No. 3.7e-165;
  Matches  650;  Conservative    0;  Mismatches   18;  Indels    2;  Gaps    2;

Qy        416 gccttggttgaatagaagtcaaacagtagtggaagagtatttggctttcttggtaatct 475
Db          1 GCCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTGGTAATCT 60

Qy        476 tgtatcagcacagactgtttttcctcagaccgtgtctcagcatgattgcttcccatttgt 535
Db         61 TGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGT 120

Qy        536 gcctcccgagtgatcattaaggaaggcgatgtagatgtttcagattctgatgatgaaga 595
Db        121 GCCTCCCCGAGTGATCATTAAGGAAGGCGATGTACATGTTTCAGATTCTGATGATGAAGA 180

Qy        596 tgataatcttcctgcaaattttgacacatgtcacagagccttgcaaataatagcaagata 655
Db        181 TGATAATCTTCCTGCAAATTTTGACACATGTCACAGAGCCTTGCAAATAATAGCAAGATA 240

Qy        656 tgtaccatcgacaccgtggtttctcatgccaatactggtggaaaaatttccatttgttcg 715
Db        241 TGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTCG 300

Qy        716 aaaatcagagagaacactggaatgttacgttcataacttactaaggattagtgtatattt 775
Db        301 AAAATCAGAGAGAACACTGGAATGTTACGTTCATAACTTACTAAGGATTAGTGTATATTT 360
```

```
Qy     776 tccaaccttgaggcatgaaattctggagcttattattgaaaaactactcaagttggatgt 835
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TCCAACCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAAACTACTCAAGTTGGATGT 420

Qy     836 gaatgcatcccggcagggtattgaagatgctgaagaaacagcaactcaaacttgtggtgg 895
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GAATGCATCCCGGCAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGGTGG 480

Qy     896 gacagattccacggaaggattgtttaatatggatgaagatgaagaaactgaacatgaaac 955
           ||||||||||||||||||| ||||||||||||||||||||| |||||||||||   |||
Db     481 GACAGATTCCACGGAAGGATCGTTTAATATGGATGAAGATGACGAAACTGAACACTCAAC 540

Qy     956 aaaggctgg-tcctgaacggctcgaccagatggtgcatcctgtagccgagcgcctggaca 1014
           | ||||||  ||||||||||||||||||||||||||||||||||| ||| ||||||||||
Db     541 ACAGGCTGNTTCCTGAACGGCTCGACCAGATGGTGCATCCTGTAGTCGACCGCCTGGACA 600

Qy    1015 tcctgatgtctttcggtttggcctacatgcaggatgtctg-tatgtagatggtaagggtg 1073
           |||||||||||||| | |||| |||||||| |||||||||| |||||||||||||||| ||
Db     601 TCCTGATGTCTTTCGGTTTGGCCTACATGCAGGATGTCTG-TATGTAGATGGTAAGGGTG 659

Qy    1074 ataacggcaa 1084
           ||| || | |
Db     660 ATACCGCCCA 669


RESULT   8
BI464721
LOCUS       BI464721                 736 bp    mRNA    linear   EST 21-AUG-2001
DEFINITION  603202341F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5268011 5',
            mRNA sequence.
ACCESSION   BI464721
VERSION     BI464721.1  GI:15255364
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 736)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11676  row: b  column: 12
            High quality sequence stop: 725.
FEATURES             Location/Qualifiers
     source          1. .736
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5268011"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      206 a    143 c    178 g    209 t
ORIGIN
```

```
  Query Match           29.7%;  Score 614.6;  DB 10;  Length 736;
  Best Local Similarity   95.4%;  Pred. No. 3.2e-164;
  Matches  699;  Conservative    0;  Mismatches   24;  Indels   10;  Gaps    6;

Qy      34 qgcctgaggcgcagcggtcgcgttagttcggcccaatggcggcaccgctgcttcacacgc 93
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6 GGCTTGAGGCGCAGCGGTCGCGTTAGTTCGGCCCAATGGCGGCACCGCTGCTTCACACGC 65

Qy      94 gtttgccgggagatgcggccgcttcgtcctctgcagttaagaagctgggcgcgtcgaggga 153
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66 GTTTGCCGGGAGATGCGGCCGCTTCGTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGA 125

Qy     154 ctgggatttcaaaatatgcatgcattagaagaatgacttttttcaatttctccccccaaagaaaaa 213
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     126 CTGGGATTTCAAATATGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAA 185

Qy     214 ctgttcggtttggtggaactgtgacagaagtcttgctgaaagtaca--aaaagggtgaaaac 271
           ||||||||||||||||||||||||||||||||||||||||||||||  || ||||||||||
Db     186 CTGTTCGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAACAACAGGGTGAAAC 245

Qy     272 aaatgacttttgaagttgttgaagaaccagctgttagatccagacataaaaggatgaccagat 331
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     246 AAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAAAGGATGACCAGAT 305

Qy     332 catcaactggctgctagaattccgttcttctatcatgtacttgacaaaagactttgagca 391
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     306 CATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATGTACTTGACAAAAGACTTTGAGCA 365

Qy     392 acttatcagtattatattaagattgccttggttgaatagaagtcaaacagtagtggaaga 451
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     366 ACTTATCAGTATTATATTAAGATTGCCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGA 425

Qy     452 gtatttggctttttcttggtaatatcttgtatcagcacagactgttttcctcagaccgtgtct 511
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     426 GTATTTGGCTTTTCTTGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCT 485

Qy     512 cagcatgattgcttcccatttgtgcctccccg-agtgatcattaaggaaggcgatgt-a 569
           |||||||||||| ||||||||||||||||||| |||||||||||||||||||||||| |
Db     486 CAGCATGATTGGTTCCCATTTGTGCCTCCCCGAAGTGATCATTAAGGAAGGCGATGTAA 545

Qy     570 gatgtttcagatttctgatgatgaagatgataatcttcctgcaaatttgaacaca--- tg 625
           |||||||||||||||||||||||||||||||||||||    || ||||  |||    ||
Db     546 GATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCGTGCAAATTTGTGACAACAGTG 605

Qy     626 tcacagagccttg-aaataatagcaagatatgtaccatcgacacccgtggtttctcatggc 684
           ||||||||||||| |||||||||||||||||||||||||||||||||| ||||||||||||
Db     606 TCACAGAGCCTTG-AAATAATAGCAAGATATGTACCATCGACA-CGTGGTTTCTCATGGC 663

Qy     685 aatactggtggaaaaatttccatttgttcgaaaatcagaagaaacactggaatgttacgt 744
           |||||||||||||||| |  ||  |  |||||||||||||||  | ||||||||||||
Db     664 AATACTGGTGGAAAAATTCCATTTGTTCGAAAAATCAGAGAGAACATGGAAATGTTACGT 723

Qy     745 tcataacttacta 758
           |||| | ||||||
Db     724 TCATTAATTACTA 736


RESULT   9
BG167061
LOCUS       BG167061                 990 bp    mRNA    linear   EST 06-FEB 2001
DEFINITION  602344740F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4455009 5',
            mRNA sequence.
ACCESSION   BG167061
VERSION     BG167061.1  GI:12673764
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 990)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
```

```
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10247  row: o  column: 10
            High quality sequence stop: 665.
FEATURES             Location/Qualifiers
     source          1..990
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4455009"
                     /clone_lib "NIH_MGC_89"
                     /tissue_type "hypernephroma, cell line"
                     /lab_host "DH10B (phage resistant)"
                     /note "Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.8 kb. Library enriched for
                     full length clones and constructed by Life Technologies.
                     Note: this is a NIH MGC Library."
BASE COUNT       304 a    275 c    219 g    192 t
ORIGIN


  Query Match             28.2%;  Score 582.2;  DB 10;  Length 990;
  Best Local Similarity   98.5%;  Pred. No. 6.8e-155;
  Matches  598;  Conservative    0;  Mismatches    8;  Indels    1;  Gaps    1;

Qy    1463 cacctttgttttttagacacaaqcaqctttqaqcqqaaacctqaaaqqaqqtttqcaqta 1522
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CACCTTTGTTTTTACACACAAGCAGCTTTGAGCGGAAACCTGAAAGAAGGTTTGCAGTA 60

Qy    1523 tcttcaqaqtctqaatttqaqcqqatagtqatqaqccaqctaaa-tccctqaaaqattt 1581
           |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db      61 TCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAAGTCCCCTGAAGATTT 120

Qy    1582 qcctqcctcaqtqtttaactttttqctqcaatcacaaataaqtaccaqctcqtcttct 1641
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCCTGCCCTCAGTGGTTAACTTTTTTGCTGCAATCACAAATAAGTACCAGCTCGTCTTCT 180

Qy    1642 qctaracccatcattqaaaqqaaacaatcqccaqatqctqccaqtcattaqqaqtaccqctq 1701
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GCTACACCATCATTGAGAGGAACAATCGCCAGATGCTGCCAGTCATTAGGAGTACCGCTG 240

Qy    1702 qaqqaqactcaqtqcaqatctqcacaaaccccqctqqacacctcttccccctttqatcccct 1761
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTCTTCCCCTTTGATCCCT 300

Qy    1762 qtqtqctqaaqaqqtcaaaqaaattcattqatcctatttatcaqqtatqaqaaqacatqa 1821
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTTATCAGGTGTGGGAAGACATGA 360

Qy    1822 qtqctqaaqaqctacaqqaqttcaaqaaaccccatqaaaaaqqacatacqtqaaqatqaaq 1881
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAAGGACATACGTGAAGATGAAG 420

Qy    1882 atqatqacttttctqaaaqqcqaaqtqcccccaqaatqatacccqtqattqqqatcacaccaa 1941
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ATGATGACTTTCTGAAAGGCGAAGTGCCCCCAGAATGATACCCGTGATTGGGATCACACCAA 480

Qy    1942 qctcctttqacacqcatttccqaaqtccttcaaqtaqtqtqqqctccccaccccqtqttqt 2001
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GCTCCTTTGACACGCATTTCCGAAGTCCTTCAAGTAGTGTGGGCTCCCCAACCGTGTAGT 540

Qy    2002 acatqcaacccaqtcccctctqacqqcaqaaatttqtqactqaqatqtqacatttqqqat 2061
           |||||||| |||||||||||||||||||||| |||||||||||||||||||  || ||
Db     541 ACATGCAAGGCAGTCCCCTCTGACGGCAGAAATATGTGACTGACATGTGACATAAGGCAT 600

Qy    2062 tcccccat 2068
           ||||||||
Db     601 TCCCCAT 607


RESULT  10
BI761160
LOCUS       BI761160                 828 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  604043657F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5183895 5',
            mRNA sequence.
ACCESSION   BI761160
VERSION     BI761160.1  GI:15752738
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 828)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11459  row: a  column: 16
            High quality sequence stop: 822.
FEATURES             Location/Qualifiers
     source          1..828
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5183895"
                     /clone_lib "NIH_MGC_116"
                     /lab_host "DH10B"
                     /note "Organ: pooled colon, kidney, stomach; Vector:
                     pCMV SPORT6; Site 1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 3 colons, age 26 yo male, 49 yo
                     female, 71 yo male colon; 46 yo male kidney, and pool of 2
                     stomachs, 62 yo male and 70 yo female.  Library is
                     oligo-dT primed and  directionally cloned (EcoRV site is
                     destroyed upon cloning).  Average insert size 1.4 kb,
                     insert size range 1-3 kb.  Library is normalized and
                     enriched for full-length clones and was constructed by C.
                     Gruber  (Invitrogen).  Research Genetics tracking code
                     023. Note: this is a NIH_MGC Library."
BASE COUNT       203 a    215 c    207 g    203 t
ORIGIN


  Query Match             26.8%;  Score 553.4;  DB 10;  Length 828;
  Best Local Similarity   97.9%;  Pred. No. 1.1e-146;
  Matches  571;  Conservative    0;  Mismatches   11;  Indels    1;  Gaps    1;

Qy    1156 qqqattcqccaqaaqcatttttqqaaacatctctqqaaaaaattqccaqqacccaaatcatcc 1255
           |||||||||||||||||||||||||| |||||||||||||||| ||||||||||||
Db     161 GGGATTCGCCAGAGGCATTTTTGGAACATCTTTGGAAAAAATTGCAGGATCCAAGTAATCC 220

Qy    1256 tqccatcatcaqqcaqqctqctqqaaattatattqqaaqctttttqqcaaaqaqctaaatt 1315
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 TGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAGAGCTAAATT 280

Qy    1316 tattcctcttattactqtaaaatcatqcctaqatctttqqttaactqqctqcaratata 1375
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     281 TATTCCTCTTATTACTGTAAAACCATGCCTAGATCTTTGGTTAACTGGCTGCACATATA 340

Qy    1376 rcttaataaccaqqattcqqqaacaaaqqcattctqcqatqttqctctccatqqaccatt 1435
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     341 TCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCAATGTTGCTCTCCATGGACCATT 400
```

```
Qy   1436 ttactcagcctgccaagctgtgttctacaccttttgttttttagacacaagcagctttttgag 1495
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    401 TTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTTAGACACAAGCAGCTTTTTGAG 460

Qy   1496 cggaaacctgaaagaaggtttgcagtatcttcagagtctgaaattttgagcggatagtgat 1555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    461 CGGAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGAT 520

Qy   1556 gagccagctaaatcccctgaagatttgcctgccctcagtggttaacttttttgctgcaat 1615
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    521 GAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGCAAT 580

Qy   1616 cacaaataagtaccagctcgtcttctgctacaccatcattgagaggaacaatcgccagat 1675
          |||| ||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    581 CACACATAAGTACCAGCTCGTCTTCTGCTACACCATCATCGAGAGGAACAATCGCCAGAT 640

Qy   1676 gctgccagtcattaggagtaccgctggaggagactcagtgcagatctgcacaaacccgct 1735
          |||||||||||||||||||| |||||||||||||||||||| |||||||||||| ||
Db    641 GCTGCCAGTCATTAGGAGTACCACTGGAGGAGACTCAGTGCAGACCTGCACAAACCCACT 700

Qy   1736 ggacaccttcttcccc-tttgatccctgtgtgctgaagagatc 1777
          ||||||||| ||||| |||||||||||||||||||| ||
Db    701 GGACACCTTCGTCCCCTTTTGATCCCTGTGTGCTGAAGAGATC 743


RESULT  11
BE870692
LOCUS       BE870692                 742 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601448528F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3852342 5',
            mRNA sequence.
ACCESSION   BE870692
VERSION     BE870692.1  GI:10319468
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 742)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9574  row: h  column: 07
            High quality sequence stop: 737.
FEATURES             Location/Qualifiers
     source          1. .742
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3852342"
                     /clone_lib="NIH_MGC_65"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.8 kb.   Library constructed by Life
                     Technologies.  "
BASE COUNT      192 a    178 c    180 g    192 t
ORIGIN


  Query Match           26.6%;  Score 550.4;  DB 10;  Length 742;
  Best Local Similarity   98.6%;  Pred. No. 7.3e-146;
  Matches  639;  Conservative    0;  Mismatches    1;  Indels    8;  Gaps    8;
```

```
Qy   1422 ctccatggaccatttttactcagcctgccaagctgtgttctacaccttttgttttttagacac 1481
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      1 CTCCATGGACCATTTTTACTCAGCCTGCCAAGCTGTGTTCTACACC-TTGTTTTTAGACAC 59

Qy   1482 aagcagctttttgagcggaaacctgaaagaaggtttgcagtatcttcagagtctgaatttt 1541
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     60 AAGCAGC-TTTGAGCGGAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAA-TTT 117

Qy   1542 gagcggatagtgatgagccagctaaatccccctgaagatttgcctgccctcagtggttaac 1601
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    118 GAGCGGATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAAC 177

Qy   1602 ttttttgctgcaatcacaaataagtaccagctcgtcttctgctacaccatcattgagagg 1661
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178 TTTTTTGCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGG 237

Qy   1662 aacaatcgccagatgctgccagtcattaggagtaccgctggaggagactcagtgcagatc 1721
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    238 AACAATCGCCAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATC 297

Qy   1722 tgcacaaacccgctggacaccttcttcccctttgatccctgtgtgctgaagagatcaaag 1781
          |||||||||:|||||||||||||||||||| |||||||||||||||||||||||:|:|.
Db    298 TGCACAAACCCGCTGGACACCTTCTTCCCC-TTGATCCCTGTGTGCTGAAGAGGTCAAAG 356

Qy   1782 aaattcattgatcctatttatcaggtatgggaaagacatgagtgctgaaagagctacaggag 1841
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    357 AAATTCATTGATCCTATTTATCAGGTGTGGGAAGACATGAGTGCTGAAGAGCTACAGGAG 416

Qy   1842 ctcaagaaacccatgaaaaagcatagtggaagatgaagatgatgactttctgaaaggc 1901
          ||||||||||:|:|||||||||||||||||||||||||||||||||||||||||||||||
Db    417 CTCAAGAAACCCATGAAAAAGGACATAGTGGAAGATGAAGATGATGACTTTCTGAAAGGC 476

Qy   1902 gaagtgcccccagaatgatacccgtgattgggatcacaccaaagctcctttgacacgca ttt 1960
          |||||||||||||||||||||| |||||||||||||||||||||:|||||||| |||
Db    477 GAAGTGCCCCCAGAATGATACCGTGA-TGGGATCACACCAAGCTCCTTTGACACGCA TTT 535

Qy   1961 ccgaaagtccttcaagtagtgtgggctcccccaccccgtgttgtacatgcaaccccagtcccct 2020
          ||||||||||||||||||||||||||||||||||||| |||||||||| ||||||||||
Db    536 CCGAAGTCCTTCAAGTAGTGTGGGCTCCCCACCCGTG-TGTACATGCAA-CCAGTCCCCT 593

Qy   2021 ctgacggcagaaatttgtgactgagatgtgacatttgggattccccat 2068
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    594 CTGACGGCAGAAATTTGTGACTGAGATGTGACATTTGGGATTCCCCAT 641


RESULT  12
AL597171
LOCUS       AL597171                 550 bp    mRNA    linear   EST 14 AUG-2001
DEFINITION  DKFZp313K0411_r1 313 (synonym: hlcc2) Homo sapiens cDNA clone
            DKFZp313K0411 5', mRNA sequence.
ACCESSION   AL597171
VERSION     AL597171.1  GI:15154983
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Koehrer,K., Beyer,A., Mewes,W., Weil,B. and Wiemann,S.
  TITLE     EST (Koehrer,K., Beyer,A., Mewes,H.W., Weil,B. and Wiemann,S.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Koehrer K
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by BMFZ (Biomedical Research Center at the Charite,
            Berlin/Germany) within the cDNA sequencing consortium of the German
            Genome Project.
            No sl sequence available.
            This clone (DKFZp313K0411) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
```

```
            Berlin Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .550
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "DKFZp313K0411"
                     /clone_lib "313 (synonym: hlcc2)"
                     /dev_stage "adult"
                     /lab_host "DH10B"
                     /note "Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA collection"
BASE COUNT       140 a    116 c    140 g    153 t      1 others
ORIGIN


  Query Match             26.5%;  Score 548;  DB 9;  Length 550;
  Best Local Similarity   99.8%;  Pred. No. 3e-145;
  Matches  548;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      24 qctqqqcatccaqqcctqaqqcqcaqcqqtcqcqttaqttcqqccaatqqcqqcaccqct 82
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 GCTGGGCATCCGGCCTGAGGCGCAGCGGTCGCGTTAGTTCGGCCCAATGGCGGCACCGCT 61

Qy      83 qcttcacacqcqtttqccqqqaqatqccqqccqcttcqtcctctqcaqttaaqaaqctqqq 142
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 GCTTCACACGCGTTTGCCGGGAGATGCCGGCCGCTTCGTCCTCTGCAGTTAAGAAGCTGGG 121

Qy     143 cqcqtcqaqqactqqqatttcaaatatqcqtqcattaqaqaatqacttttcaatttctcc 202
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 CGCGTCGAGGACTGGGATTTCAAATATGCGTGCATTAGAGAATGACTTTTTCAATTCTCC 181

Qy     203 ccccaaaaaaaaactqttcqqtttqqtqqaactqtqacaqaaqtcttqtctqaaqtacaaaaa 262
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 CCCAAGAAAAACTGTTCGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAA 241

Qy     263 qqatqaaacaaatqatcttqaqttqttqaaqaaccaqctqttaqatccaqacatqaaqqa 322
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 GGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAAAGGA 301

Qy     323 tqaccaqatcatcaactqqctqctqaqaattccqttcttctatcatqtacttqacaaaaqa 382
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 TGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATGTACTTGACAAAAGA 361

Qy     383 cttqaaqcaacttatcaqtattatattaaqattqccttqqttqaatagaaqtcaaacaqt 442
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 CTTGAGCAACTTATCAGTATTATATTAAGATTGCCTTGGTTGAATAGAAGTCAAACAGT 421

Qy     443 aqtqaaaqaqtatttqqctttttcttqqtaatctttqtatcaqcacaqactqttttcctcaq 502
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 AGTGGAAGAGTATTTGGCTTTTCTTGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAG 481

Qy     503 accqtqtctcaqcatqattqcttcccatttqtqcctccccqaqtqatcattaaqqaaqq 562
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 ACCGTGTCTCAGCATGATTGCTTCCCATTTGTGCCTCCCCGAGTGATCATTAAGGAAGG 541

Qy     563 cqatqtaqa 571
           ||||||| |
Db     542 CGATGTANA 550


RESULT  14
BI549394
LOCUS       BI549394                 727 bp    mRNA    linear   EST 05 SEP 2001
DEFINITION  604190049F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5261100 5',
            mRNA sequence.
ACCESSION   BI549394
VERSION     BI549394.1  GI:15446706
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 727)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11658  row: b  column: 14
            High quality sequence stop: 715.
FEATURES             Location/Qualifiers
     source          1. .727
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:5261100"
                     /clone_lib "NIH_MGC_95"
                     /tissue_type "hippocampus"
                     /lab_host "DH10B"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI XhoI (qtcgaq
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN 3',
                     size-selected for  average insert size 2.5 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for  full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health).  Note: this is a NIH_MGC Library."
BASE COUNT       182 a    191 c    171 g    183 t
ORIGIN


  Query Match             26.5%;  Score 547.2;  DB 10;  Length 727;
  Best Local Similarity   97.6%;  Pred. No. 5.9e-145;
  Matches  566;  Conservative    0;  Mismatches   13;  Indels    1;  Gaps    1;

Qy    1156 qpqattcqccaqaqcqcattttttqqaaccatcttctqcqaaaaaaattqcaqqaccccaaqtaatcc 1255
           |||||||||||||:|||||||||||||||||||| |||||||||||||||||| ||||||||||
Db     106 GGGATTCGCAGAGGCATTTTTGGAACATCTTTGGAAAAAATTGCAGGATCCAAGTAATCC 165

Qy    1256 tqccatcatcaqqcaqqctqctqqaaaattatattqqaaqctttttqqccaaqaqctaaatt 1415
           ||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db     166 TGCCATCATCAGCAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAGAGCTAAATT 225

Qy    1316 tattcctcttattactqtaaaatcatqcctaqatctttqqttaactqqctqccacatata 1375
           |||||:|||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     226 TATTCCTCTTATTACTGTAAAACCATGCCTAGATCTTTGGTTAACTGGCTGCACATATA 285

Qy    1376 cttaataaccaqqattcqqqaacaaaqqcattctqcqatqttqctctccatqqaccatt 1435
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     286 CCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCAATGTTGCTCTCCATGGACCATT 345

Qy    1436 ttactcaqcctqccaaqctqtqttctacaccttqttttqaqcacaaqcaqctttqaq 1495
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     346 TACTCAGCCTGCCAAGCTGTGTTCTACACCTTAGTTTTGAGCACAAGCAGCTTTGAG 405

Qy    1496 cqqaaacctqaaaqaaqqtttqcaqtatcttcaqaqtctqaaatttqaqcqqatatqtqat 1555
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     406 CGGAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTGAGCGGATATGTGAT 465

Qy    1556 qaqccaqctaaatccctqaaqatttqcctqccctcaqtqqttaacttttttqctqcaat 1615
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     466  AGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTAGCTGCAAT 525

Qy    1616 cacaaatqaqtqccaqctcqtcttctqctacaccatcatttqaqaqqaacaatcqccaqat 1675
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     526  CACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATCGAGAGGAACAATCGCCAGAT 585

Qy    1676 qctqccaqtcatttaqqaqtaccqctqqaqqaqactcaqtqcaqatctqcacaaaccqct 1735
```

```
          |||||||||||||||||||| |||||||||||||||||||| ||||||||| || ||
Db    586 GCTGCCAGTCATTAGGAGTACCACTGGAGGAGACTCAGTGCAGACCTGCACAAAACCACT 645

Qy   1746 qgacacctt-cttccccttgatccctgtgtgctgaaqaq 1774
          ||||| ||| | ||||||||||||||||||||||||||||
Db    646 GGACAACTTACGTCCCCTTTGATCCCTGTGTGCTGAAGAG 685


RESULT  14
AW408066
LOCUS       AW408066                 563 bp    mRNA    linear   EST 16-FEB-2000
DEFINITION  U1-HF-BM0-adu-b-10-0-U1.r1 NIH_MGC_38 Homo sapiens cDNA clone
            IMAGE:3062754 5', mRNA sequence.
ACCESSION   AW408066
VERSION     AW408066.1  GI:6927123
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 563)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Eco RI site shown at the beginning of the sequence.
             Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: M.B. Soares Lab
             cDNA Library Arrayed by: M.B. Soares Lab
             DNA Sequencing by: M.B. Soares Lab
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .563
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3062754"
                     /clone_lib="NIH_MGC_38"
                     /tissue_type="lymph"
                     /cell_type="germinal center B cells"
                     /cell_line="MGC85"
                     /lab_host="DH10B (LT1)"
                     /note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
                     Constructed from size fractionated cytoplasmic mRNA
                     (2.5-3.5kb). Directionally cloned. Cells provided by Louis
                     M. Staudt, Ph.D. Library preparation by Maria de Fatima
                     Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT      151 a    122 c    119 g    171 t
ORIGIN


  Query Match             26.2%;  Score 542.4;  DB 9;  Length 563;
  Best Local Similarity   99.6%;  Pred. No. 1.2e-143;
  Matches  554;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy   1054 qctatgtaqatggtaaggttgataacggcaaaacaaaggatctatatcgcgacctgataa 1113
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8 GCTATGTAGATGGTAAGGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAA 67

Qy   1114 acatctttgacaaactcctgttgcccacccatgcctcctgccatgtacagttttcatgt 1173
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     68 ACATCTTTGACAAACTCCTGTTGCCCACCCATGCCTCCTGCCATGTACAGTTTTTCATGT 127

Qy   1174 tttacctctgtagtttcaaattgggattcgcagaggcatttttggaacatctctggaaaa 1233
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    128 TTTACCTCTGTAGTTTCAAATTGGGATTCGCAGAGGCATTTTTGGAACATCTCTGGAAAA 187

Qy   1234 aattgcaggacccaagtaatcctgccatcatcaggcaggctgctggaaattatattggaa 1293
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    188 AATTGCAGGACCCAAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAA 247

Qy   1294 qctttttqqcaaqaqctaaatttattcctcttattactgtaaaatcatgcctagatcttt 1353
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    248 GCTTTTTGGCAAGAGCTAAATTTATTCCTCTTATTACTGTAAAATCATGCCTAGATCTTT 307

Qy   1354 tqgttaactggctgcacatatacctttaataaccaqgattcqqqaacaaaqqcattctgcq 1413
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    308 TGGTTAACTGGCTGCACATATACCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCG 367

Qy   1414 atgttgctctccatgqaccattttactcaqcctgccaaqctgtgttctacaccttttatt 1473
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    368 ATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTTATT 427

Qy   1474 ttaqacacaaqcaqctttqaqccqaaacctqaaaqaaqgtttgca-gtatcttcaqaqt 1532
          ||||||||||||||||||||||||||||||||||||||||||| ||| ||||| |||
Db    428 TTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGTTTGCAGGTACCTTCAGAGT 487

Qy   1533 ctgaattttgaqcqqataqtqatgaqccaqctaaatcccctqaaqatttqcctgcctca 1592
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    488 CTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCTCA 547

Qy   1593 gtqqttaacttttttq 1608
          ||||||||||||||||
Db    548 GTGGTTAACTTTTTTG 563


RESULT  15
AW958173
LOCUS       AW958173                 564 bp    mRNA    linear   EST 01-JUN-2000
DEFINITION  EST370243 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW958173
VERSION     AW958173.1  GI:8147856
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 117
            Seq primer: Reverse.
FEATURES             Location/Qualifiers
     source          1. .564
                     /organism "Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGE"
                     /note="Vector: pBluescriptSKm"
BASE COUNT      157 a    108 c    130 g    168 t      1 others
ORIGIN


  Query Match             26.0%;  Score 536.8;  DB 9;  Length 564;
  Best Local Similarity   99.1%;  Pred. No. 4.8e-142;
  Matches  560;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    2;

Qy     82 tqcttcacacqcqtttqccqqqaqatqcqqccqcttcqtcctctqcagttaaqaaqctqq 141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGCTTCACACGCGTTTGCCGGGAGATGCGGCCGCTTCGTCCTCTGCAGTTAAGAAGCTGG 60

Qy    142 acqcqtcqaqqactqqqatttcaaaatatqcqtqcattaqaqaaatqacttttttcaattctc 201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCGCGTCGAGGACTGGGATTTCAAATATGCGTGCATTAGAGAATGACTTTTTCAATTCTC 120
```

```
Qy      202 ccccaaqaaaaaactqttcaqttttqatqqaactqtqacaqaaqtcttqctqaaqtacaaaa 261
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 CCCCAACAAAAACTGTTCCGTTTGTTGGAACTGTCACAGAAGTCTTGCTGAAGTACAAAA 180

Qy      262 atqqtqaaacaaatqactttqaqttqttqaaqaaccaqctqttaqatccaqacataaaqq 321
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 AGGCTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAAAGG 240

Qy      322 atqaccaqatcatcaactqctqctaqaattccqttcttctatcatqtacttqacaaaaq 381
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 ATGACCAGATCATCAACTGCTGCTAGAATTCCGTTCTTCTATCATGTACTTGACAAAAG 300

Qy      382 actttqaqcaacttatcaqtattatattaaqattqccttqqttqaataqaaqtcaaacaq 441
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 ACTTTGAGCAACTTATCAGTATTATATTAAGATTGCCTTGGTTGAATAGAAGTCAAACAG 360

Qy      442 taqtqqaaqaqtatttqqctttcttqqtaatcttqtatcaqcacaqactqttttcctca 501
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 TAGTGGAAGAGTATTTGGCTTTTCTTGGTAATCTTGTATCAGCACAGACTGTTTTCCTCA 420

Qy      502 qaccqtqtctcaqcatqattqcttcccatttqtqcctccccqaqtqatcatt aaqqaa 560
            |||||||||||||||||||||| |||||||||||||||||||||||||||| ||||||
Db      421 GACCGTGTCTCAGCATGATTGCTTTCCATTTGTGCCTCCCCGAGTGATCATTAAAGGAA 480

Qy      561 qqcqatqtaqatqtttcaqattctqatqatqaaqatqataatcttcctqcaaatttqaac 620
            |||||||||||||||||||||||||||||||||||||||| ||||| ||||||||
Db      481 GGCGATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTT CTGCANATTTTGAC 539

Qy      621 acatqtcacaqaqcctqcaaataa 645
            |||||||||||||| ||||||||
Db      540 ACATGTCACAGAGCCCTGCAAATAA 564
```

Search completed: July 30, 2002, 03:09:51
Job time: 3761 sec

```
OM of: US-09-932-678-2 to: N_Geneseq_032802:*   out_format : pfs

Date: Jul 30, 2002  6:29 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=x11
 -Q=/cgn2_1/USPTO_spool/US09932678/runat_29072002_155124_17977/app_query.fasta_1.717
 -DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=std.rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09932678_@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-932-678-2
Query length: 651
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 214.350000



score_list:
Sequence            Strd Orig    ZScore    EScore Len  ! Documentation  ..
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA98384 + 3391.00 6461.01  0.0       20
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC76937 + 2162.00 4110.18  1.4e-220  14
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS01562 + 1973.00 3747.96  2.0e-200   1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS44584 + 1763.00 3344.92  5.8e-178   1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS87641 + 1313.00 2473.70  1.9e-129   3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS92254 + 1205.00 2266.71  6.6e-118   3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS87636 + 1181.00 2234.23  4.2e-116   1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA83038 + 1124.00 2120.70  8.9e-110   1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS92255 + 1074.00 2018.95  4.1e-104   2
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS92253 +  882.00 1664.46  2.3e-84    7
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS87643 +  851.00 1599.31  9.8e-81    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS80955 -  831.00 1568.74  4.9e-79    6
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX51663 +  712.00 1345.50  1.3e-66   43
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS91991 +  689.00 1298.48  5.6e-64    5
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS80954 +  610.00 1151.21  8.9e-56    3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS87635 +  610.00 1151.21  8.9e-56    3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS91986 +  610.00 1151.21  8.9e-56    3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS92246 +  610.00 1151.21  8.9e-56    3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS92247 +  577.00 1085.25  4.2e-52    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC77533 +  556.00 1049.22  4.3e-50   34
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS92248 -  550.50 1032.56  3.6e-49    5
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAV90248 +  505.00  951.77  1.1e-44   34
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS73660 +  487.00  891.51  2.6e-41    2
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS91992 +  487.00  891.51  2.6e-41    2
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL03627 +  400.00  733.26  1.7e-32    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK68676 +  398.00  699.80  1.2e-30    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK68677 +  392.00  680.54  1.5e-29    3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK85168 +  392.00  680.54  1.5e-29    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC39853 +  387.50  704.40  6.9e-31   22
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC50235 +  387.00  703.45  7.8e-31   22
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS92251 +  385.50  714.81  1.8e-31    6
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK68674 +  372.00  657.73  2.7e-28    9
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK68678 +  372.00  646.53  1.2e-27    2
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK85169 +  372.00  646.53  1.2e-27    2
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL03626 +  366.50  651.98  5.7e-28    6
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC19091 +  364.00  687.15  6.3e-30   21
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC77798 +  322.00  579.50  6.2e-24   21
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS87640 +  284.50  522.00  1.0e-20    6
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS92252 -  258.50  469.18  8.7e-18    8
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS91907 -  257.00  460.41  2.7e-17    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS70319 +  257.00  450.56  9.6e-17    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL18528 +  229.00  402.28  4.6e-14    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL18529 +  209.50  382.77  5.7e-14    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS79656 +  206.50  356.69  1.6e-11    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:AAT20548 +  187.00  342.77  9.6e-11    1

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA98384

seq_documentation_block:
ID   AAA98384 standard; cDNA; 2040 BP.
XX
AC   AAA98384;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human RNA polymerase I transcription factor TIF-1A cDNA.
XX
KW   RNA polymerase I transcription factor TIF-1A; antitumor; treatment;
KW   antiproliferative; cell proliferation; cancer; tissue regeneration; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055316-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-DE000767.
XX
PR   17-MAR-1999;  99DE-1011992.
XX
PA   (DEKR ) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI   Grummt I, Vingron M;
XX
DR   WPI; 2000-587527/55.
DR   P-PSDB; AAB10936.
XX
PT   New DNA encoding the transcription factor TIF-1A, useful for preventing
PT   or treating diseases associated with abnormal cell proliferation,
PT   particularly tumors
XX
PS   Claim 1; Fig 2; 38pp; German.
XX
CC   This invention describes a novel DNA sequence (I) that encodes the RNA
CC   polymerase I transcription factor TIF-1A which has antitumor,
CC   antiproliferative and proliferation-inducing activity. The invention also
CC   describes (1) DNA (Ia) encoding a protein (II) with the biological
CC   activity of TIF-1A; (2) a ribozyme (R) corresponding to (I) or (Ia) and
CC   able to bind specifically to, and cleave, its transcribed RNA so as to
CC   reduce or inhibit synthesis of the corresponding protein; (3) an
CC   antisense RNA (AS) with binding properties similar to R; (4) an
CC   expression vector that contains (I), (Ia) or sequences that encode R or
CC   AS; (5) host cells containing the vectors of (4); (6) TIF-1A or (II)
CC   encoded by (I) or (Ia); (7) preparation of TIF-1A or (II) by culturing
CC   cells of (6); (8) ligands that bind to TIF-1A or (II); (9) antagonists
CC   that weaken or block the activity of TIF-1A or (II); (10) a diagnostic
CC   method for detecting abnormal TIF-1A expression; and (11) kit for
CC   carrying out the method in (10). (I), and similar sequences that encode
CC   proteins with equivalent activity, expression vectors containing them, or
CC   their expression products are used to treat or prevent disorders
CC   associated with reduced cellular proliferation, to stimulate cellular
CC   proliferation, and to promote tissue regeneration, e.g. after injury or
CC   radiation therapy. Ribozymes, antisense sequences directed against (I),
CC   also ligands and antagonists of TIF-1A, are used to treat or prevent
CC   disorders associated with excessive cellular proliferation and to inhibit
CC   proliferation, especially in treatment of cancers. (I) and specific
CC   ligands for TIF-1A (particularly antibodies (Ab)) are also useful for
CC   diagnosis of altered TIF-1A expression by (in)direct determination of the
CC   concentration, length and/or sequence of TIF-1A or its mRNA, e.g. for
CC   detecting mutations. Ab can also be used for immunoprecipitation of
CC   TIF-1A and for isolation of related sequences from cDNA expression
CC   libraries. (I) allows recombinant production of TIF-1A in sufficient
CC   quantities for therapeutic use.
```

```
XX
SQ   Sequence 2040 BP; 563 A; 433 C; 474 G; 568 T; 2 other;


alignment scores:
            Quality: 3391.00              Length:    651
              Ratio: 5.217                  Gaps:      0
 Percent Similarity: 99.846   Percent Identity: 99.693

alignment block:
 US-09-932-678-2 x AAA98484    ..

 Align Seq 1/1   to: AAA98484   from: 1   to: 2040

       1 MetAlaAlaProLeuLeuHisThrArqLeuProGlyAspAlaAlaAlaSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      40 ATGGCGGCACCGCTGCTTCACACGGGTTTGCCGGGAGATGCGGCCGCTTC 89

      17 rSerSerAlaValLysLysLeuGlyAlaSerArqThrGlyIleSerAsnM 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      90 CTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATA 139

      34 etArqAlaLeuGluAsnAspPhePheAsnSerProProArqLysThrVal 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     140 TGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTT 189

      51 ArqPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGlyGl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     190 CGGTTTGGTGGAACTGTGACAGAAGTCTTGCTCAAGTACAAAAAGGGTGA 239

      67 uThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     240 AACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAA 289

      84 ysAspAspGlnIleIleAsnTrpLeuLeuGluPheArqSerSerIleMet 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     290 AGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATG 339

     101 TyrLeuThrLysAspPheGluGluLeuIleSerIleIleLeuArqLeuPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     340 TACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCC 389

     117 oTrpLeuAsnArqSerGlnThrValValGluGluTyrLeuAlaPheLeuG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     390 TTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTG 439

     134 lyAsnLeuValSerAlaGluThrValPheLeuArqProCysLeuSerMet 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     440 GTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATG 489

     151 IleAlaSerHisPheValProProArqValIleIleLysGluGlyAspVa 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     490 ATTGCTTCCCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCGATGT 539

     167 lAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPheA 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     540 AGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCTGCAAATTTTG 589

     184 spThrCysHisArqAlaLeuGluIleIleAlaArqTyrValProSerThr 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     590 ACACATGTCACAGAGCCTGCAAATAATAGCAAGATATGTACCATCGACA 639

     201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArqLy 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     640 CCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTCGAAA 689

     217 sSerGluArqThrLeuGluCysTyrValHisAsnLeuLeuArqIleSerV 234
         |||||||||||||||||||||||||||||||    ||||||||||||||||
     690 ATCAGAGAGAACACTGGAATGTTACGTTCATNACTTACTAAGGATTAGTG 739

     234 alTyrPheProThrLeuArqHisGluIleLeuGluLeuIleIleGluLys 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     740 TATATTTTCCAACCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAAA 789
```

```
     251 LeuLeuLysLeuAspValAsnAlaSerArqGlnGlyIleGluAspAlaGl 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     790 CTACTCAAGTTGGATGTGAATGCATCCCGGCAGGGTATTGAAGATGCTGA 839

     267 uGluThrAlaThrGluThrCysGlyGlyThrAspSerThrGluGlyLeuP 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     840 AGAAACACCAACTCAAACTTGTGGTGGGACAGATTCCACGGAAGGATTGT 889

     284 heAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyPro 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     890 TTAATATGGATGAAGATGAAGAAACTGAACATGAAACAAAGGCTGGTCCT 939

     301 GluArqLeuAspGluMetValHisProValAlaGluArqLeuAspIleLe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     940 GAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCGCCTGGACATCCT 989

     317 uMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyL 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     990 GATGTCTTTGGTTTGTCCTACATGAAGGATGTCTGCTATGTAGATGGTA 1039

     334 ysValAspAsnGlyLysThrLysAspLeuTyrArqAspLeuIleAsnIle 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1040 AGGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATC 1089

     351 PheAspLysLeuLeuProThrHisAlaSerCysHisValGlnPhePh 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1090 TTTGACAAACTCCTGTTGCCCACCCATGCCTCCTGCCATGTACAGTTTTT 1139

     367 eMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuG 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1140 CATGTTTTACCTCTGTAGTTTCAAATTGGGATTCGCAGAGGCATTTTTGG 1189

     384 luHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArq 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1190 AACATCTCTGGAAAAAATTGCAGGACCCAAGTAATCCTGCCATCATCAGG 1239

     401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArqAlaLysPheIl 417
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1240 CAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAGAGCTAAATTTAT 1289

     417 eProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuH 434
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1290 TCCCTCATTATTACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGC 1339

     434 isIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspVal 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1340 ACATATACCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTT 1389

     451 AlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPh 467
         ||||||||||||||||||||||||||||||||||||||||||||||:::||
    1390 GCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACNCCTT 1439

     467 eValPheArqHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuG 484
         |:|||||||||||||||||||||||||||||||||||||||||||||||||
    1440 TGTTTTTAGA'ACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGTTTGC 1489

     484 luTyrLeuGluSerLeuAsnPheGluArqIleValMetSerGluLeuAsn 500
         |:|||||||||||||||||||||||||||||||||||||||||||||||||
    1490 ACTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAAT 1539

     501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleTh 517
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1540 CCCCTGAAGATTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGCAATCAC 1589

     517 rAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArqAsnAsnA 534
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1590 AAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAATC 1639

     534 rgGlnMetLeuProValIleArqSerThrAlaGlyGlyAspSerValGln 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1640 GCCAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAG 1689
```

```
     551 IleCysThrAsnProLeuAspThrPhePheProPheAspProCysValIle 567
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1690 ATCTGCACAAACCCGCTGGACACCTTCTTCCCCTTTGATCCCTGTGTGCT 1739

     567 uLysArgSerLysLysPheIleAspProIleTyrGlnValTrpGluAspM 584
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1740 GAAGAGGTCAAAGAAATTCATTGATCCTATTTATCAGGTGTGGGAAGACA 1789

     584 etSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1790 TGAGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAAGGACATA 1839

     601 ValGluAspGluAspAspAspPheLeuLysGlyGluValProGlnAsnAs 617
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1840 GTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCAGAATGA 1889

     617 pThrValIleGlyIleThrProSerSerPheAspThrHisPheArgSerP 634
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1890 TACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAGTC 1939

     634 roSerSerSerValGlySerProProValLeuTyrMetGlnProSerPro 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1940 CTTCAAGTAGTGTGGGCTCCCCACCCGTGTTGTACATGCAACCCAGTCCC 1989

     651 Leu 651
         |||
    1990 CTC 1992

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC76937

seq_documentation_block:
ID   AAC76937 standard; cDNA; 1418 BP.
XX
AC   AAC76937;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF2492 polynucleotide sequence SEQ ID NO:4983.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB42728.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease
XX
PS   Claim 5; Page 4157-4158; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to innibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 1418 BP; 422 A; 275 C; 318 G; 403 T; 0 other;


alignment_scores:
            Quality: 2152.00              Length:    422
              Ratio:  5.185                 Gaps:      1
 Percent Similarity:  98.815   Percent Identity: 97.867

alignment_block:
 US-09-932-678-2 x AAC76937   ..

 Align seq 1/1  to: AAC76937  from: 1  to: 1418

     202 TrpPheLeuMetProIleLeuValGluLysPheProPheValArgLysSe 218
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       3 TGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTCGAAAATC 52

     218 rGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerValT 235
         |'||||||||||||||||||||||||||||||||||||||||||||||||||
      53 AGAGAGAACA'TGCAATGTTACGTTCATAACTTACTAAGGATTAGTGTAT 102

     235 yrPheProThr.LeuArgHisGluIleLeuGluLeuIleIleGluLysIe 251
         | ||||||||| ||||||||||||||||||||||||||||||||||||||
     103 ATTTTCCAACCCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAAACT 152

     251 uLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluG 268
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     153 ACTCAAGTTGGATGTGAATGCATCCCGGCAGGGTATTGAAGATGCTGAAG 202

     268 luThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPhe 284
         |||||||||||||||||  ||||||||||||||||||||||||||||||
     203 AAACAGCAACTCAAACTTTGGTGGGACAGATCCACGGAAGGATTGTTT 252

     285 AsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyProGl 301
         |.||||||||||||||||||||||||||||||||||||||||||||||||
     253 AATATGGATGAAGATGAAGAAACTGAACATGAAACAAAGGCTGGTCCTGA 302

     301 uArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuM 318
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     303 ACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCGCCTGGACATCCTGA 352

     319 etSerLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyLys 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     353 TGTCTTTGGTTTGTCCTACATGAAGGATGTCTGCTATGTAGATGGTAAG 402
```

```
 345 ValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIlePh 351
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 403 GTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATCTT 452

 351 eAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheM 368
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 453 TGACAAACTCCTGTTGCCCACCCATGCCTCCTGCCATGTACAGTTTTTCA 502

 368 etPheTyrLeuCysSerPheLysLeuGlyPheAlaGlnAlaPheLeuGlu 384
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 503 TGTTTTACCTCTGTAGTTTCAAATTGGGATTCGCAGAGGCATTTTTGGAA 552

 385 HisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGl 401
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 553 CATCTCTGGAAAAAATTGCAGGACCCAAGTAATCCTGCCATCATCAGGCA 602

 401 nAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleP 418
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 603 GGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAGAGCTAAATTTATTC 652

 418 roLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHis 434
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 653 CTCTTATTACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCAC 702

 435 IleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAl 451
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 703 ATATACCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTTGC 752

 451 aLeuHisGlyProPheTyrSerAlaCysGluAlaValPheTyrThrPheV 468
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 753 TCTCCATGGACCATTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTG 802

 468 alPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGln 484
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 804 TTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGTTTGCAG 852

 485 TyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnPr 501
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 853 TATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAATCC 902

 501 oLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrA 518
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 903 CCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGCAATCACAA 952

 518 snLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArg 534
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 953 ATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAATCGC 1002

 535 GlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIl 551
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1003 CAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGAC 1052

 551 eCysThrAsnProLeuAspIleThrPhePheProPheAspProCysValLeuL 568
      ||||||||||||||||||||||||||||||||||||||||||||||||||
1054 CTGCACAAACCCACTGGACACCTTCTTCCCCTTTGATCCCTGTGTGCTGA 1102

 568 ysArgSerLysLysPheIleAspProIleTyrGlnValTrpGluAspMet 584
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1103 AGAGGTCAAAGAAATTCATTGATCCTATTTATCAGGTGTGGGAAGACATG 1152

 585 SerAlaGluGluLeuGluGluPheLysLysProMetLysLysAspIleVa 601
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1153 AGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAAGGACATAGT 1202

 601 lGluAspGluAspAspAspPheLeuLysGlyGluValProGlnAsnAspI 618
     ||||||||||||||||||||||||||||||:::|||||||:::   :
1203 GGAAGATGAAGATGATGACTTTCTGAAAGGCGAAATTCCCCAGAAATTAG 1252

 618 hrValIleGlyIle 622
     ::|||   |||:::
1253 TAGTAAGTGGGGTC 1266
```

```
seq name: /SIDS1/acqdata/geneseq/geneseqn emb1/NA2001A.DAT:AAS01562

seq_documentation block:
ID   AAS01562 standard; cDNA; 1423 BP.
XX
AC   AAS01562;
XX
DT   18-JUL-2001  (first entry)
XX
DE   Human secretory molecule cDNA sptm #52.
XX
KW   Human; secretory molecule; sptm; SPTM; library screening; gene therapy;
KW   cell signalling; cell proliferative disorder; atherosclerosis; cancer;
KW   immune system disorder; AIDS; neurological disorder; Alzheimer's disease;
KW   nervous system disease; mental retardation; developmental disorder;
KW   neuromuscular disorder; microarray; Incyte ID number 4287452dec; ss.
XX
OS   Homo sapiens.
XX
PN   WO200123558-A2.
XX
PD   05-APR-2001.
XX
PF   19-SEP-2000; 2000WO-US25610.
XX
PR   28-SEP-1999;   99US-0156624.
PR   28-SEP-1999;   99US-0156625.
PR   02-DEC-1999;   99US-0168611.
PR   02-DEC-1999;   99US-0168613.
PR   02-DEC-1999;   99US-0168614.
XX
PA   (INCY ) INCYTE GENOMICS INC.
XX
PI   Hodgson DM,  Lincoln SE,  Russo ED,  Spiro PA,  Banville SC;
PI   Bratcher SR,  Dufour GE,  Cohen BJ,  Rosen BR,  Shah P,  Chalup MS;
PI   Billman JL,  Jones AL,  Yu JY,  Greenawalt LB,  Panzer SR;
PI   Roseberry AM,  Wright RJ,  Chen W,  Liu TF,  Yap PE,  Stockdreher PK;
PI   Anshey S,  Fong WT;
XX
DR   WPI; 2001-258134/26.
XX
PT   New secretory polynucleotides (sptm) and the polypeptides they encode,
PT   for use in inducing antibodies and screening libraries of compounds
XX
PS   Claim 1; Page 152; 161pp; English.
XX
CC   The present sequence for human secretory molecule cDNA sptm #52
CC   (Incyte ID number 4287452dec) is 1 of 63 novel sptm cDNA sequences
CC   (AAS01511-AAS01574) which encode for the secretory polypeptides SPTM. The
CC   sptm polynucleotides are useful for screening a compound for
CC   effectiveness in altering expression of a target polynucleotide, where
CC   the target polynucleotide comprises sptm. Sptm is also useful in a method
CC   for assessing the toxicity of a test compound. Sptm and its fragments or
CC   complementary sequences, may be used to identify the presence of and/or
CC   determine the degree of similarity between two nucleic acid sequences.
CC   Sptm can also be used for a variety of diagnostic and therapeutic
CC   purposes, e.g. diagnosing a particular condition, disease or disorder
CC   associated with cell signalling, such conditions include cell
CC   proliferative disorders such as atherosclerosis, and cancers including
CC   leukaemia, an immune system disorder e.g. acquired immunodeficiency
CC   syndrome (AIDS), a neurological disorder such as epilepsy or Alzheimer's
CC   disease, nutritional and metabolic disease of the nervous system, mental
CC   retardation and other developmental disorders, and muscular dystrophy and
CC   other neuromuscular disorders. Sptm can also be used to design probes
CC   useful in diagnostic assays, which may be used to monitor the progress of
CC   conditions or disorders associated with abnormal levels of expression of
CC   sptm. In addition sptm encoding SPTM may be used for somatic or germline
CC   gene therapy, for inducing antibodies, or in microarrays.
XX
SQ   Sequence 1423 BP; 417 A; 247 C; 420 G; 439 T; 0 other;


alignment_scores:
```

```
           Quality: 1973.00          Length:     488
             Ratio:   4.995            Gaps:       2
 Percent Similarity:  80.943   Percent Identity: 80.533

alignment_block:
 US-09-932-678-2 x AAS01562    ..

 Align seq 1/1  to: AAS01562  from: 1  to: 1423

      31 IleSerAsnMetArgAlaLeuGluAsnAspPhePheAsnSerProProAr 47
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     194 ATTTCAAATATGCGTGCATTAGAGAATGATTTTTTCAATTCTCCCCCAAG 243

      47 gLysThrValArgPheGlyGlyThrValThrGluValLeuLeuLysTyrL 64
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     244 AAAAACTGTTCGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACA 293

      64 ysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     294 AAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGAT 343

      81 ProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSe 97
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     344 CCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTC 393

      97 rSerIleMetTyrLeuThrLysAspPheGluGlnLeuIleSerIleIleL 114
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     394 TTCTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATAT 443

     114 euArgLeuProTrpLeuAsnArgSerGlnThrValValGluGluTyrLeu 130
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     444 TAAGATTGCCTTGGTTGAATAGAAGTCAAACACTAGTGGAAGAGTATTTG 493

     131 AlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCy 147
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     494 GCTTTTCTTGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTG 543

     147 sLeuSerMetIleAlaSerHisPheValProProArgValIleIleLysG 164
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     544 TCTCAGCATGATTGCTTCCCATTTTGTGCCTCCCCGAGTGATCATTAAGG 593

     164 luGlyAspValAspValSerAspSerAspAspGluAspAspAsnLeuPro 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     594 AAGGCGATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCT 643

     181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrVa 197
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     644 GCAAATTTTGACACATGTCACAGAGCCTTGCAAATAATAGCAAGATATGT 693

     197 lProSerThrProTrpPheLeuMetProIleLeuValGluLysPheProP 214
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     694 ACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCAT 743

     214 heValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeu 230
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     744 TTGTTCGAAAATCAGAGAGAACACTGGAATGTTACGTTCATAACTTACTA 793

     231 ArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLeuIl 247
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     794 AGGATTAGTGTATATTTTCCAACCTTGAGGCATGAAATTCTGGAGCTTAT 843

     247 eIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleG 264
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     844 TATTGAAAAACTACTCAAGTTGGATGTGAATGCATCCCGGCAGGGTATTG 893

     264 luAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr 280
         |||||||||||||||||||||:::||||||||||||||||||||||||||
     894 AAGATGCTGAAGAAACAGCAAATCAAACTTGTGGTGGGACAGATTCCACG 943

     281 GluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLy 297
         ||||||||||||||||||
     944 GAAGGATTGTTTAATATG................................ 961

     297 sAlaGlyProGluArgLeuAspGlnMetValHisProValAlaGluArgL 314

     961 .................................................. 961

     314 euAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyr 330

     961 .................................................. 961

     331 ValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLe 347

     961 .................................................. 961

     347 uIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysHisV 364

     961 .................................................. 961

     364 aiGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGlu 380
                                               ||||||||||||
     962 ......................................GGATTCGCAGAG 973

     381 Ala.PheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProA 397
         ||| ||||||||||||||||||::::||||||||||||||||||||||||
     974 GCCATTTTTGGAACATCTTTGGAAAAACTTGCAGGATCCAAGTAATCCTG 1023

     397 laIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArg 413
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1024 CCATCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAGA 1073

     414 AlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuVa 430
         |||||||||||    ||||||||||||||    ||||||||||||||||
    1074 GCTAAATTTATTTCTCTTATTACTGTAAAACCATGCCTAGATCTTTTGGT 1123

     430 lAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaP 447
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1124 TAACTGGCTGCACATATACCTTAATAACCAGGATTCGGGAACAAAGGCAT 1173

     447 heCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaVal 463
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1174 TCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTG 1223

     464 PheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLy 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1224 TTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAA 1273

     480 sGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetS 497
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1274 AGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGA 1323

     497 erGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPhePhe 513
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1324 GCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTT 1373

     514 AlaAlaIleThr 517
         ||||||||||||
    1374 GCTGCAATCACA 1385

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAS44584

seq_documentation_block:
ID   AAS44584 standard; DNA; 1498 BP.
XX
AC   AAS44584;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human full-length polynucleotide sequence #9.
XX
KW   Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW   mouse: Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW   cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW   nervous system disorder; inflammatory disorder; cell differentiation; ds;
KW   angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW   genetic disorder; bone regeneration; tendon; ligament; tissue repair;
```

```
KW   cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;
KW   antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW   neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;
KW   immunostimulant; analgesic; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200164864-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04926.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
PR   17-JUN-2000; 2000US-0597707.
PR   14-JUL-2000; 2000US-0616807.
PR   19-SEP-2000; 2000US-0664641.
XX
PA   (HYSE ) HYSEQ INC.
XX
PI   Tang Y.T.,  Liu C.,  Zhou P.,  Asundi V.,  Zhang J.,  Zhao Q.A.,  Ren F.;
PI   Xue A.J.,  Yang Y.,  Wehrman T.,  Wang J.,  Ma Y.,  Wang D.,  Chen R.,  Xu C.;
PI   Drmanac R.;
XX
DR   WPI; 2001-589862/66.
DR   P-PSDB; AAU27684.
XX
PT   Novel polypeptides and nucleic acids obtained from cDNA libraries
PT   prepared from various human tissues, for diagnosis, treatment of
PT   cancer, neurological, inflammatory disorders and for use in arrays for
PT   detection -
XX
PS   Claim 1; SEQ ID No 9; 153pp; English.
XX
CC   Sequences AAS44576-AAS44919 represent full-length polynucleotides and
CC   contig polynucleotides encoding polypeptides of the invention. The DNA
CC   and protein sequences are useful for the treatment, diagnosis and
CC   prevention of various types of disorder in a mammalian subject such as a
CC   human, dog, monkey, mouse, hamster or rat. The disorders include cancers
CC   such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such
CC   as multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC   diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC   disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC   chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC   Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC   disease, ischaemia reperfusion injury, shock, sepsis and inflammatory
CC   bowel disease. The sequences exhibit activity relating to angiogenesis,
CC   cell proliferation, cell differentiation, stem cell growth factor,
CC   activin or inhibin. Therefore, they can be used to manipulate stem cells
CC   in culture to give rise to neuroepithelial cells that can be used to
CC   augment or replace cells damaged by illness, accidental damage or genetic
CC   disorders. The sequences may also be used for regeneration of bone,
CC   cartilage, tendons and ligaments and in tissue repair and burn healing.
CC   Note: Some sequences for this patent did not form part of the printed
CC   specification, but were obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1498 BP; 417 A; 332 C; 425 G; 424 T; 0 other;


Alignment scores:
            Quality: 1763.00              Length:     447
              Ratio:    5.155               Gaps:       0
 Percent Similarity:   98.559   Percent Identity:  97.695

Alignment block:
 US-09-932-678-2 x AAS44584    ..

 Align Seq 1/1   to: AAS44584   from: 1   to: 1498

     223 GluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLe 239
         ||| ||||||||||||||||  ||||||||||||||||||||||||||||

       4 GAAAGTTACGTTCATAACTTACAGAGGATTAGTGTATATTTTCCAACCTG 52

     239 uArgHisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspV 256
         |||||||||||:||||:::|||||||:::|||||||||||||||||||||
      53 .AGGCATGAAATTCTGAGGCTTATTATGGAAAAACTACTCAAGTTGGATG 102

     255 alAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThrGln 272
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     103 TGAATGCATCCCGGCAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAA 152

     273 IhrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGluAs 289
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     153 ACTTGTGGTGGGACAGATTCCACGGAAGGATTGTTTAATATGGATGAAGA 202

     289 peIleGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGlnM 306
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     203 TGAAGAAACTGAACATGAAACAAAGGCTGGTCCTGAACGGCTCGACCAGA 252

     306 etValHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeu 322
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     253 TGGTGCATCCTGTAGCCGAGCGCCTGGACATCCTGATGTCTTTGGTTTTG 302

     323 SerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLy 339
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     303 TCCTACATGAAGGATGTCTGCTATGTAGATGGTAAGGTTGATAACGGCAA 352

     339 sThrLysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuL 356
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     353 AACAAAGGATCTATATCGCGACCTGATAAACATCTTTGACAAACTCCTGT 402

     359 euProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCys 372
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     403 TCCCACCCATGCCTCCTGCCATGTACAGTTTTTCATGTTTTACCTCTGT 452

     373 SerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLy 389
         |||||||||| |||||||||||||||||||||||||||||||||||||||
     453 AGTTTCAAATTGGCATTCGCAGAGGCATTTTTGGAACATCTCTGGAAAAA 502

     389 sLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnT 406
         ||||||||||| |||||||||||||||||||||||||||||||||||||||
     503 ATTGCAGGACCCAAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATT 552

     423 yrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrVal 422
         |||||||||||||||||||||||||||||||||||||||||||||||||
     553 ATATTGGAAGCTTTTTGGCAAGAGCTAAATTTATATCCTCTTATTACTGTA 602

     423 LysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAs 439
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     603 AAATCATGCCTAGATCTTTGGTTAACTGGCTGCACATATACCTTAATAA 652

     439 nGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProP 456
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      65 C AGGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATGGACCAT 702

     459 hrTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLys 472
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      70 T TACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAG 752

     473 GlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLe 489
         |.||||||||||||||||||||||||||||||||||||||||||||||||
      75 CaGCTTTTGAGCGCAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCT 802

     489 sAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCysL 506
         |.|||||||||||||||||||||||||||||||||||||||||||||||||
      80 GAATTTGAGCGGATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCC 852

     506 euProSerValValAsnPhePheAlaAlaIleThrAsnLysTyrGluLeu 522
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      85 TTCCCTCAGTGGTTAACTTTTTTGCTGCAATCACAAATAAGTACCAGCTC 902

     523 ValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProVa 539
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      90 GTCTTCTGCTACACCATCATCGAGAGGAACAATCGCCAGATGCTGCCAGT 952
```

```
     539 IIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProL 556
         ||||||||||||||:::|||||||||||||||||||   ||||||||||||
     953 CATTAGGAGTACCACTGGAGGAGACTCAGTGCAGACCTGCACAAACCCAC 1002

     556 euAspThrPhePheProPheAspProCysValLeuLysArq 569
         ||||||||||||   |||||||||||||||||||||||||||
    1003 TGGACACCTTCGTCCCCTTTGATCCCTGTGTGCTGAAGAGA 1043

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA20018.DAT:AAS87641

seq_documentation_block:
ID   AAS87641 standard; cDNA; 3169 BP.
XX
AC   AAS87641;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #23445.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG23454.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 23445; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3169 BP; 812 A; 841 C; 779 G; 737 T; 0 other;
```

```
alignment_scores:
            Quality: 1313.00              Length:    498
              Ratio:  3.979                Gaps:     10
 Percent Similarity:  66.265   Percent Identity: 58.835

alignment_block:
 US-09-932-678-2 x AAS87641   ..

 Align seq 1/1  to: AAS87641  from: 1  to: 3169

       1 MetAlaAlaProLeuLeuHisThrArqLeuProGlyAspAlaAlaAlaSe 17
         |||||||||||||||||||||||||e|||||||||||||||||||||||||
       1 ATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGCAGATGCGGCCGCTTC 50

      17 rSerSerAlaValLysLysLeuGlyAlaSerArqThrGlyIleSerAsnM 34
         |   ||||||||||||   ||||||||||||||||||||||||||||||||
      51 GCCCTCTGCAGTCAAGATGCTGGGCGCGTCGAGGACTGGGATTTCAAATA 100

      34 etArqAlaLeuGluAsnAspPhePheAsnSerProProArqLysThrVal 50
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     101 TGCGTGCATTAGACAATGATTTTTTCAATTCTCCCCCAAGAAAAACTGTT 150

      51 ArqPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGlyGl 67
         :::|||||||||||||||||||||||||||||||:||||||   |||||
     151 CAGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAACGGGTGA 200

      67 uThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleL 84
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     201 AACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAA 250

      84 ys......AspAspGlnIleIleAsnTrpLeuLeuGluPheArqSerSer 98
         ||       :::          |||   |||   ||::::
     251 AGGTGGTGTGCCCCGCCATGCCGATGATCCTGAGGGAGGGGAGGGCCATT 300

      99 Ile.............................MetTyrLeuThrLysAs 105
         :::                           :::   |||   :::::
     301 GTGGGCGAGTGTGC'TTTCATCCCCTTCGTGGCTTTCAGTTGTGCTCAAA 350

     105 pPheGluGlnLeuIleSerIleIleLeuArqLeuProTrpLeuAsnArqS 122
         :   :::   |||         :::   :::|||||   ::::::|||
     351 C'AGAGTGGCCTCAGCCCGAAGGTTGCTAAGCTGCCTGCGGTGGGTCGG. 399

     122 erGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSer 138

     399 ..................................................... 399

     139 AlaGlnThrValPheLeuArqProCysLeuSerMetIleAlaSerHisPh 155
               |||       ||||||      |||             |||
     400 ......CCCGTGGCCTGCCGGCCCACATACAGCCCCAGCCTCTGCCAC.. 441

     155 eValProProArqValIleIleLysGluGlyAspValAspValSerAspS 172

     441 ..................................................... 441

     172 erAspAspGluAspAspAsnLeuProAla.AsnPheAspThrCysHisAr 188
                 :::   :::   ||||||  :::|||          |||
     442 ..........AACCCCCAACGCCCAGCGCAGCTTCTGGCTCACTCAT.. 478

     188 qAlaLeuGlnIleIleAlaArqTyrValProSerThrProTrpPheLeuM 205
                 ::::::::::|||            ||||||||||||||||
     479 .........CTGCTTTGCAGG............ACACCGTGGTTTCTCA 506

     205 etProIleLeuValGluLysPheProPheValArqLysSerGluArqThr 221
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     507 TCCCAATACTGGTGGAAAAATTTCCATTTGTTCGAAAATCAGAGAGAACA 556

     222 LeuGluCysTyrValHisAsnLeuLeuArqIleSerValTyrPheProTh 238
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     557 CTGGAATGTTACGTTCATAACTTACTAAGGATTAGTGTATATTTTCCAAC 606

     238 rLeuArqHisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuA 255
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     607 CTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAAACTACTCAAGTTGG 656
```

```
     255 spValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThr 271
         ||||||||||||||||||||||||||||||||||||||||||||||||||:::
     657 ATGTGAATGCATCCCGGCAGGGTATTGAAGATGCTGAAGAAACAGCAAAT 706

     272 GlnPheCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGl 288
         |||||||||||||||||||||||||||||||||||||||    ||||||   ::
     707 CAAACTTCTGCGTGGCACAGATTCCACGGAAGGATTGTCTAATATGGTTAG 756

     288 nAspGluGluThrGluHisGluPheLysAlaGlyProGluArgLeuAspG 305
         :
     757 C..................................................... 757

     305 lnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuVal 321
                                      ::::::||||||:::::
     758 ..........................AGTTTATTAATGAAAGTGGAG 778

     322 LeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGl 338
         ::::::::::::
     779 ATGAAGTTTATC..................................... 790

     338 yLysThrLysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuL 355
                                  :::|||:::   :::   ::::||:
     791 ..........................ATAATCAAACGGTGGAAACAGCTAG 815

     355 euLeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeu 371
         ::|||
     816 TGCTG..................................................... 820

     372 CysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLy 388
              :::   :::|||||||||||||||||||||||||   ||||||||
     821 ......CTCATCTTTGGATTCGCCAGAGGCATTTTTGGAACCTCTTTGGAA 864

     388 sLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyA 405
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     865 AAAATTGCAGGATCCAACTAATCCTGCCATCATCAGGCAGGCTGCTGGAA 914

     405 snTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThr 421
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     915 ATTATATTGCAAGCTTTTTGCCAAGAGCTAAATTTATTCCTCTTATTACT 964

     422 ValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAs 438
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     965 GTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAA 1014

     438 nAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyP 455
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1015 CAACCAGGATTCGGCAACAAAGGCATTCTGCGATGTTGCTCTCCATGGAC 1064

     455 roPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHis 471
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1065 CATTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACAC 1114

     472 LysGluLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyr 485
         ||||||||||||||||||||||||||||||:::   :::
    1115 AAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGCGTCTGGTTC 1156

seq name: /SIDS1/gcgdata/geneseq/geneseqn_embl/NA2001B.DAT:AAS92254

seq documentation block:
ID   AAS92254 standard; cDNA; 4175 BP.
XX
AC   AAS92254;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #28058.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0644167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639462/74.
DR   P-PSDB; AAG28067.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID NO 28058; 104pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4175 BP; 811 A; 849 C; 779 G; 736 T; 0 other;


alignment_scores:
            Quality: 1205.00              Length:    504
              Ratio:  3.685                Gaps:     16
 Percent Similarity:  64.881   Percent Identity: 57.143

alignment_block:
 US-09-932-678-2 x AAS92254   ..

 Align seq 1/1  to: AAS92254  from: 1  to: 4175

       1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGCAGATGCGGCCGCTTC 50

      17 rSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnM 34
         |   ||||||||||||   ||||||||||||||||||||||||||||||||
      51 GCCCTCTGCAGTCAAGATGCTGGGCGCGTCGAGGACTGGGATTTCAAATA 100

      34 etArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrVal 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 TCCGTGCATTAGAGAATGATTTTTTCAATTCTCCCCCAAGAAAAACTGTT 150

      51 ArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGlyGl 67
         ::|||||||||||||||||||||||||||||||||||||||   |||||
     151 CAGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGA 200
```

```
 67 uThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleL 84
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
201 AACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAA 250

 84 ys......AspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSer 98
    ||          :::         |||   |||   |||:::
251 AGGTGGTGTGCCCCGCCATGCCGATGATCCTGAGGGAGGGGAGGGCCATT 300

 99 Ile.............................MetTyrLeuThrLysAs 105
    :::                           :::   |||   :::::
301 GTGGGCGAGTGTGCCTTTCATCCCCTTCGTGGCTTTCAGTTGTGCTCAAA 350

105 pPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgS 122
    :   :::   |||       :::   :::|||||   :::::::||
351 CCAGAGTGGCCTCAGCCCGAAGGTTGCTAAGCTGCCTGCGGTGGGICGG. 399

122 erGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSer 138

399 ..................................................... 399

139 AlaGlnThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPh 155
            |||     ||||||     |||           |||
400 ......CCCGTGGCCTGCCGGCCCACATACAGCCCCAGCCTCTGCCAC.. 441

155 eValProProArgValIleIleLysGluGlyAspValAspValSerAspS 172

441 ..................................................... 441

172 erAspAspGluAspAspAsnLeuProAla.AsnPheAspThrCysHisAr 188
            :::   :::   ||||||  :::|||        |||
442 ..........AACCCCCAACGCCCAGCGCAGCTTCTGGCTCACTCAT.. 478

188 qAlaLeuGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuM 205
            ::::::::::|||           ||||||||||||||||
479 .........CTGCTTTGCAGG...........ACACCGTGGTTTCTCA 506

205 etProIleLeuValGluLysPheProPheValArgLysSerGluArgThr 221
    ||||||||||||||||||||||||||||||||||||||||||||||||||
507 TGCCAATACTGGTGGAAAAATTTCCATTTGTTCGAAAATCAGAGAGAACA 556

222 LeuGluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProTh 238
    ||||||||||||||||||||||||||||||||||||||||||||||||||
557 CTGGAATGTTACGTTCATAACTTACTAAGCATTAGTGTATATTTTCCAAC 606

238 rLeuArgHisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuA 255
    ||||||||||||||||||||||||||||||||||||||||||||||||||
607 CTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAAACTACTCAAGTTGG 656

255 spValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThr 271
    ||||||||||||||||||||||||||||||||||||||||||||||||:::
657 ATGTGAATGCATCCCGGCAGGGTATTGAAGATGCTGAAGAAACAGCAAAT 706

272 GlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGl 288
    ||||||||||||||||||||||||||||||||||   ||||||   ::
707 CAAACTTGTGGTGGGACAGATTCCACGGAAGGATTGTGTAATATGGTTAG 756

288 uAspGluGluThrGluHisGlnThrLysAlaGlyProGluArgLeuAspG 305
    :
757 C.................................................. 757

305 lnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuVal 321
                              ::::::||||||:::::
758 ............................AGTTTATTAATGAAAGTGGAG 778

322 LeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGl 338
    ::::::::::::
779 ATGAAGTTTATC..................................... 790

338 yLysThrLysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuL 355
                              :::|||:::   :::   :::|||:
791 .........................ATAATCAAAGGGTGGAAACAGCTAG 815

355 euLeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeu 371
    ::||||
816 TGCTG.............................................. 820

372 CysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLy 388
            :::   :::|||||||||||||||||||||||||   ||||||||
821 ......CTCATCTTTGGATTCGCAGAGGCATTTTTGGAACCTCTTTGGAA 864

388 sLysLeuGln.AspProSerAsnProAlaIleIleArgGlnAlaAlaGly 404
    |||||||||| |||||||||||||||||||||||||||||||||||||||
865 AAAACTGCAGGGATCCAAGTAATCCTGCCATCATCAGGCAGGCTGCTGGA 914

405 AsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleTh 421
    :::|||||:::|||||||||||||||||||   ||||||   ||||||||
915 CATTATATTGCAAGCTTTTTGGCAAGAGCTACATTTATTTCTCTTATTAC 964

421 rValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuA 438
    |||||||   ||||||||||||||||||||||||||||||||||||||||
965 TGTAAAACCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTA 1014

438 snAsnGlnAspSer.GlyThrLysAlaPheCysAspValAla.LeuHisG 454
    |||||||||||||| ||||||||||||||||||||||||||| |||||||
1015 ATAACCAGGATCCGGGAACAAAGGCATTCTGCGATGTTGCCTCTCCATG 1064

454 lv.ProPheTyrSerAlaCys.GlnAlaValPheTyrThrPheValPheA 470
    || |||||||:|||||||||| ||||||||||||||||||||||||||||
1065 GACCCATTTTACTCAGCCTGCCCAAGCTGTGTTCTACACCTTTGTTTTTA 1114

470 rg.HisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyr 485
    || ||||||||||||||||||||||||||||||||||||:::   :::
1115 GAACACAAGCAGCTTTGAGCGGAAACCTGAAAGAAGGCCTCTGGTTC 1162
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS87636

seq_documentation_block:

```
ID   AAS87636 standard; cDNA; 1017 BP.
XX
AC   AAS87636;
XX
DT   15-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #23440.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   24-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG23449.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 23440; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
```

```
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1017 BP; 301 A; 214 C; 240 G; 272 T; 0 other;


alignment scores:
            Quality: 1181.00              Length:    452
              Ratio: 4.687                  Gaps:      2
 Percent Similarity: 55.752   Percent Identity: 55.088

alignment block:
 US-09-932-678-2 x AAS87636   ..

 Align seq 1/1  to: AAS87636  from: 1  to: 1017

       30 sGlyIleSerAsnMetArgAlaLeuGluAsnAspPhePheAsnSerProPr  46
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        7 GGCATTTCAAATATGCGTGCATTAGAGAATGATTTTTTCAATTCTCCCCC  56

       46 oArgLysThrValArgPheGlyGlyThrValThrGlnValLeuLeuLysT  63
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       57 AAGAAAAACTCTTCGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGT  106

       64 yrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeu  79
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      107 ACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTA  156

       80 AspProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheAr  96
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      157 GATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAGAATTCCG  206

       96 qSerSerIleMetTyrLeuThrLysAspPheGluGlnLeuIleSerIleI  113
          |||||||:::||||||||||||||||||||||||||||||||||||||||
      207 TTCTTCTGTCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTA  256

      113 leLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGluTyr  129
          ||
      257 TA..................................................  258

      130 LeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgPr  146

      258 ..................................................  258

      146 oCysLeuSerMetIleAlaSerHisPheValProProArgValIleIleL  163

      258 ..................................................  258

      163 ysGluGlyAspValAspValSerAspSerAspAspGluAspAspAsnLeu  179

      258 ..................................................  258

      180 ProAlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTy  196

      258 ..................................................  258

      196 rValProSerThrProTrpPheLeuMetProIleLeuValGluLysPheP  213

      258 ..................................................  258

      213 roPheValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeu  229
                                    |||||||||||||||||||||||||
      259 ..........................TTGGAATGTTACGTTCATAACTTA  282

      230 LeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLe  246
          |||||||||||||||||||||||||||||||||||||||||||||||||||
      283 CTAAGGATTAGTGTATATTTTCCAACCTTGAGGCATGAAATTCTGGAGCT  332

      246 uIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyI  263
          |||||||||||||||||||||||||||||||||||||||||||||||||||
      333 TATTATTGAAAAACTACTCAAGCTGGATGTGAATGCATCCCGGCAGGGTA  382

      263 leGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSer  279
          |||||||||||||||||||||||:::|||||||||||||||||||||||||
      383 TTGAACATGCTGAAGAAACAGCAAATCAAACTTGTGGTGGGACAGATTCC  432

      280 ThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluTh  296
          ||||||||||||||||||||
      433 ACGGAAGGATTGTTTAATATG..............................  454

      296 rLysAlaGlyProGluArgLeuAspGlnMetValHisProValAlaGluA  313

      454 ..................................................  454

      313 rgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCys  329

      454 ..................................................  454

      330 TyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAs  346

      454 ..................................................  454

      346 pLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysH  363

      454 ..................................................  454

      363 isValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAla  379
                                                 |||||||||
      454 ..........................................GGATTCGCA  462

      380 GluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnPr  396
          |||||||||||||||||||||||||:::||||||||||||||||||||||
      463 GAGGCATTTTTGGAACATCTTTGGAAAAACTTGCAGGATCCAAGTAATCC  512

      396 oAlaIleIleArgGluAlaAlaGlyAsnTyrIleGlySerPheLeuAlaA  413
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      513 TGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAA  562

      413 rgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeu  429
          |||||||||||||||  ||||||||||||||  ||||||||||||||||
      563 GAGCTAAATTTATTTCTCTTATTACTGTAAAACCATGCCTAGATCTTTTG  612

      430 ValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAl  446
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      613 GTTAACTGGCTGCACATATACTTAATAACCAGGATTCGGGAACAAAGGC  662

      446 aPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaV  463
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      663 ATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTG  712

      463 alPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeu  479
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      713 TTTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTG  762

      480 LysGlu 481
          ||||||
      763 AAAGAA 768

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAB48038
```

```
seq_documentation_block:
ID   ABA83038 standard; DNA; 1461 BP.
XX
AC   ABA83038;
XX
DT   05-FEB-2002  (first entry)
XX
DE   Human transcription factor TREX-65 coding sequence.
XX
KW   Human; transcription factor; TREX; cell proliferative disease;
KW   autoimmune disease; inflammation; neurological disease;
KW   developmental disorder; cancer; AIDS; infection; cytostatic; anti-HIV;
KW   neuroprotective; antiinflammatory; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200172777-A2.
XX
PD   04-OCT-2001.
XX
PF   13-MAR-2001; 2001WO-US08117.
XX
PR   13-MAR-2000; 2000US-0188986.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Hillman JL,  Baughn MR,  Yue H,  Lal P,  Lu DAM,  Patterson C;
PI   Azimzai Y,  Bandman O,  Tang YT,  Mathur P,  Shah P,  Au-Young J;
PI   Reddy R;
XX
DR   WPI; 2001-570896/64.
DR   P-PSDB; ABB50214.
XX
PT   Novel transcription factor polypeptides, used to treat diseases
PT   associated with altered activity and expression of TREX, and to screen
PT   for agents capable of modulating its activity -
XX
PS   Claim 11; Page 299; 327pp; English.
XX
CC   The present sequence is the coding sequence for a human transcription
CC   factor. The transcription factor and its coding sequence are useful in
CC   the diagnosis, treatment and prevention of diseases associated with
CC   altered expression of the transcription factor e.g. cell proliferative,
CC   autoimmune/inflammatory, neurological and developmental disorders. A
CC   number of specific disorders/diseases are given in the specification,
CC   including: arteriosclerosis, cirrhosis, hepatitis, cancers, AIDS,
CC   allergies, anaemia, asthma, autoimmune thyroiditis, bronchitis, atopic
CC   dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,
CC   Grave's disease, multiple sclerosis, osteoarthritis, pancreatitis,
CC   psoriasis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative
CC   colitis, uveitis, Alzheimer's disease, Huntington's disease, Parkinson's
CC   disease, stroke, and viral, bacterial, fungal and protozoal infections.
XX
SQ   Sequence 1461 BP; 413 A; 335 C; 298 G; 415 T; 0 other;


alignment_scores:
            Quality: 1124.00              Length:    486
              Ratio:  3.986                Gaps:      8
 Percent Similarity:  58.025   Percent Identity: 51.646

alignment_block:
 US-09-932-678-2 x ABA83038   ..

 Align seq 1/1  to: ABA83038  from: 1  to: 1461

     195 ArqTyrValProSerThrProTrpPheLeuMetProIleLeuValGluLy 211
         ::::||          :::|||||||||||||||||||||||||||||||
     263 AAATACAAAATTAGGGCACCGTGGTTTCTCATGCCAATACTGCTGCAAAA 312

     211 sPheProPheValArqLysSerGluArqThrLeuGluCysTyrValHisA 228
         |||||||||||||||||||||||||||||||||||||||||||||||||

     313 ATTTCCATTTGTTCGAAAATCAGAGAGAACACTGGAATGTTACGTTCATA 362

     228 snLeuLeuArqIleSerValTyrPheProThrLeuArqHisGluIleLeu 244
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     363 ACTTACTAAGGATTAGTGTATATTTTCCAACCTTGAGGCATGAAATTCTG 412

     245 GluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArqGl 261
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     413 GAGCTTATTATTGAAAAACTACTCAAGCTGGATGTGAATGCATCCGGCA 462

     261 nGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrA 278
         ||||||||||:|||||||||||||||||:::|||||||||||||||||||
     463 GGGTATTGAAGATGCTGAAGAAACAGCAAATCAAACTTGTGGTGGGACAG 512

     278 spSerThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHis 294
         ||||||||||||||||||||||||||
     513 ATTCCACGGAAGGATTGTTTAATATG..................... 538

     295 GluThrLysAlaGlyProGluArqLeuAspGlnMetValHisProValAl 311

     538 ..................................................... 538

     311 aGluArqLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspV 328

     536 ..................................................... 538

     328 alCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyr 344

     538 ..................................................... 538

     345 ArqAspLeuIleAsnIlePheAspLysLeuLeuProThrHisAlaSe 361

     538 ..................................................... 538

     361 rCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyP 378
                                                          ||||
     539 .....................................................GGAT 542

     378 hsAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSer 394
         |:|||||||||||||||||||||||||||||||:::|||||||||||||||
     543 TGGCAGAGGCATTTTTGGAACATCTTTGGAAAAACTTGCAGGATCCAAGT 592

     395 AsnProAlaIleIleArqGlnAlaAlaGlyAsnTyrIleGlySerPheLe 411
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     593 AATCCTGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTT 642

     411 uAlaArqAlaLysPheIleProLeuIleThrValLysSerCysLeuAspL 428
         ||||||||||||||||||  ||||||||||||||||  ||||||||||
     643 GGCAAGAGCTAAATTTATTTCTCTTATTACTGTAAAACCATGCCTAGATC 692

     428 euLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThr 444
         |:||||||||||||||||||||||||||||||||||||||||||||||||
     693 TTTTGGTTAACTGGCTGCACATATACCTTAATAACCAGGATTCGGGAACA 742

     445 LysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGl 461
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     743 AAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCA 792

     461 nAlaValPheTyrThrPheValPheArqHisLysGlnLeuLeuSerGlyA 478
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     793 AGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGAGCGGAA 842

     478 snLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArqIle 494
         |||||||||||||||||||||||  ||||||||||||||||||||||||
     843 ACCTGAAAGAAGGTTTGCAGTATCCTCAGAGTCTGAATTTTGAGCGGATA 892

     495 ValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAs 511
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     893 GTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAA 942

     511 nPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrI 528
         ||||||||||||||||||:::
     943 CTTTTTTGCTGCAATCACAAAG..................... 964
```

```
     528 IeIleGluArgAsnAsnArgGluMetLeuProValIleArgSerThrAla 544
                                                  :::|||:::
     965 ..................................  ..ATGAAGACTTGT 976

     545 GlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhePr 561
         |||                        |||   ::::::   |||:::
     977 GGATATGGATGGTGGTGATGGTTGCACAACAATATCAATTTATTTTATTAC 1026

     561 oPheAspProCysValLeuLys.......................... 568
             :::|||||:::   |||
    1027 CACTGAACCGTGCACTTCAAAATGCTTAAGATGGTAAGTTTTATGTTCTG 1076

     569 ........................ArgSerLysLysPheIleAspPro 576
                                  |||::::|||:::   :::   |||
    1077 TATTTTACCATAATAAAAAAAATTGTAGAGGGAAAAACAGTCTGCCTCCA 1126

     577 IleTyrGlnValTrpGlu.....................AspMetSe 585
         :::      :::|||:::                     :::   ||
    1127 CTTTTG...ATATGGGACTGCTAACATCTTCCACCTTCCGTCTCCCCTTC 1173

     585 rAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIleValG 602
         ||||   :::|||   :::::::::|||||
    1174 TGCCCCACATCTGGCAAGCTAAGAAAGCCT.................... 1204

     602 luAspGluAspAspAspPheLeuLysGlyGlnValProGlnAsnAspThr 618
                                     :::   ::::::   |||
    1205 ..........................GCTGCTCTCTCCTCTGGCACC 1225

     619 ValIleGlyIleThrProSer.............SerPheAspThrHisPh 631
                 :::   |||:::              |||   :::||||||::
    1226 AGCTGGAAATTCATACCCAACAAGCCCTAGCCTTCCCACCAGACCCACAT 1275

     631 eArgSerProSerSerSerValGlySerProProValLeuTyrMetGlnP 648
         :   ||||||||||   |||         |||                :::|
    1276 TTCATCCCCATCCCCATCGCATCCCCATCCC..............CATC 1310

     648 roSerPro 650
         ||   |||
    1311 CCCATCCC 1318

Seq name: /SIDS1/gcgdata/geneseq/geneseqn_embl/NA20018.DAT:AAS92255

Seq documentation block:
ID   AAS92255 standard; cDNA; 2410 BP.
XX
AC   AAS92255;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #28059.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE ) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639462/74.
DR   P-PSDB; AAG28068.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID NO 28059; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2410 BP; 672 A; 437 C; 547 G; 754 T; 0 other;


alignment scores:
                Quality:  1074.00              Length:     232
                  Ratio:  4.838                  Gaps:       7
 Percent Similarity:  95.690   Percent Identity:  94.828

alignment block:
 US-09-932-678-2 x AAS92255    ..

 Align seq 1/1  to: AAS92255  from: 1  to: 2410

     427 AspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAspGlnAspSerGl 443
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 GATCTTTGGTTAACTGGCTGCACATATACCTTAATAACCAGGATTCGGG 50

     443 yThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaC 460
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 AACAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCT 100

     460 ysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSer 476
         ||||||||||||||||||||||||||||||||||||||||||||||||
     101 GTCAAGCTGTGTTCTACACCTTTGTTTTAGACACAAGCAGCTTTTGAGC 150

     477 GlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluAr 493
         ||||||||||||||||||||||||||||||||||||||||||||||||
     151 CGAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGTG 200

     493 qIleValMetSerGlnLeuAsnProLeuLysIleCys.LeuProSerVal 509
         ||||||||||||||||||||||||||||||||    ||||||||||||||
     201 GATAGTGATGAGCCAGCTAAATCCCCTGAAGATATGGCTTGCCCTCAGTC 250

      51 ValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTy 526
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GTTAACTTTTTTGCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGCTA 300

     526 rThrIleIleGluArgAsnAsnArgGlnMetLeuProValIleArgSerT 543
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 CACCATCATCGAGAGGAACAATCGCCAGATGCTGCCAGTCATTAGGAGTA 350

     543 hrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhe 559
         |:::|||||||||||||||||   ||||||||||||||||||||||||||
     351 CTACTGGAGGAGACTCAGTGCAGACCTGCACAAACCCACTGGACACCTTC 400
```

```
     560 PheProPhe.AspProCysValLeuLysArgSerLysLysPhe.IleAsp 575
         ||||||||| |||||||||||||||||||||||||||||||| ||||||
     401 TTCCCCTTTGGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCCATTGAT 450

     576 ProIle.Tyr.GlnValTrp.GluAspMetSerAlaGluGluLeuGlnGl 591
         |||||| ||| |||||||| ||||||||||||||||||||||||||||||
     451 CCTATTTTATCCAGGTATGGGGAAGACATGAGTGCTGAAGAGCTACAGGA 500

     591 uPheLysLysProMetLysLysAspIleValGluAspGluAspAspAspP 608
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 GTTCAAGAAACCCATGAAAAAGGACATAGTCGAAGATGAAGATGATGACI 550

     608 heLeuLysGlyGluValProGlnAsnAspThrValIleGlyIleThrPro 624
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 TTCTGAAAGGCGAAGTGCCCCAGAATGATACCGTGATTGGGATCACACCA 600

     625 SerSerPheAspThrHisPheArgSerProSerSerSerValGlySerPr 641
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 AGCTCCTTTGACACGCATTTCCGAAGTCCTTCAAGTAGTGTGGGCTCCCC 650

     641 oProValLeuTyr.MetGlnProSerProLeu 651
         |||||||||||| :::|||||||||||||||
     651 ACCCGTGTTGTACACTGCAACCCAGTCCCCTC 682

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS92253

seq_documentation_block:
ID   AAS92253 standard; cDNA; 776 BP.
XX
AC   AAS92253;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #28057.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG28066.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 28057; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 776 BP; 206 A; 157 C; 198 G; 215 T; 0 other;


alignment_scores:
            Quality:  882.00              Length:    196
              Ratio:  4.717                 Gaps:      5
 Percent Similarity:  95.408   Percent Identity: 94.898

alignment_block:
 US-09-932-678-2 x AAS92253

 Align seq 1/1  to: AAS92253  from: 1  to: 776

       1 MetAlaAlaProLeuLeuHisThrArg.LeuProGlyAspAlaAlaAlaS 17
         |||||||||||||||||||||||||    |||   ||||||||||||||||
     198 ATGGCGGCACCGCTGCTTCACACGTTGTTTGTCGGGAGATCCGGCCGCTT 247

      17 erSerSerAlaValLysLysLeuGlyAlaSer.ArgThrGlyIleSerAs 33
         ||||||||||||||||   |||   |||||| ||||||||||||||||||
     248 CGTCCTCTGCAGTCAAGACGCTGTTCGCGTCGTAGGACTGGGATTTCAAA 297

      33 nMetArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrV 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     298 TATGCGTGCATTAGAGAATGATTTTTTCAATTCTCCCCCAAGAAAAACTG 347

      50 alArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGly 66
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     348 TICGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGT 397

      67 GluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIl 83
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     398 GAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACAT 447

      83 eLysAspAsp.GlnIleIleAsnTrp.LeuLeuGluPheArgSerSerIl 99
         |||||||||| |||||||||||||| ||||||||||||||||||||||::
     448 AAAGGATGACCCAGATCATCAACTGGGCTGCTAGAATTCCGTTCTTCTGT 497

      99 eMetTyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgL 116
         :|||||||||||||||||||||||||||||||||||||||||||||||||
     498 CATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTGAGAT 547

     116 euPro.TrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPh 132
         |||||| ||||||||||||||||||||||||||||||||||||||||||||
     548 TGCCTTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTT 597

     132 eLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuS 149
         |||||||||||||||||| ||||||||||||||||||||||||||||||||
     598 TCTTGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCA 647

     149 erMetIleAlaSerHisPheValProProArgValIleIleLysGluGly 165
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     648 GCATGATTGCTTCCCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGC 697

     165 AspValAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAs 182
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     698 GATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCTGCAAA 747

     182 nPheAspThrCysHisArgAlaLeuGln 191
         | |||||||||||||||||||||||||
     748 TTTGACACATGTCACAGAGCCTTGCAA 775
```

```
seq_name: /SIDSI/qcqdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS87643

seq documentation block:
ID   AAS87643 standard; cDNA; 1260 BP.
XX
AC   AAS87643;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #23447.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08641.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; AWG23456.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 1; SEQ ID No 23447; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1260 BP; 344 A; 227 C; 269 G; 420 T; 0 other;


alignment scores:
            Quality:  851.00              Length:    212
              Ratio:  4.464                 Gaps:      9
 Percent Similarity:  91.981   Percent Identity: 91.038

alignment block:
 US-09-932-678-2 x AAS87643   ..

Align seq 1/1  to: AAS87643  from: 1  to: 1260

      1 MetAlaAlaProLeuLeuHisThrArg.LeuProGlyAspAlaAlaAlaS 17
        ||||||||||||||||||||||||||    |||   ||||||||||||||||
    198 ATGGCGGCAGCGCTGCTTCACACGTTGTTTGTCGGGAGATGCGGCCGCTT 247

     17 erSerSerAlaValLysLysLeuGlyAlaSer.ArgThrGlyIleSerAs 33
        |||||||||||||||||   |||   |||||| ||||||||||||||||||
    248 CGTCCTCTGCAGTCAAGACGCTGTTCGCGTCGTAGGACTGGGATTTCAAA 297

     33 nMetArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrV 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    298 TATGCGTGCATTAGAGAATGATTTTTTCAATTCTCCCCCAAGAAAAACTG 347

     50 alArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGly 66
        |||||||||||||||||||||||||||||||||||||||||||||||||
    348 TTCGGTTTGGTGCAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGT 397

     67 GluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIl 83
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    398 GAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACAT 447

     83 eLysAspAsp.GlnIleIleAsnTrp.LeuLeuGluPheArgSerSerIl 99
        ||||||||||| ||||||||||||||| ||||||||||||||||||||::
    448 AAAGGATGACCCAGATCATCAACTGGGCTGCTAGAATTCCGTTCTTCTGT 497

     99 eMetTyrLeuThr.LysAspPheGlu.GlnLeuIleSerIleIleLeuAr 115
        :|||||||||||| |||||||||||| ||||||||||||||||||||||||
    498 CATGTACTTGACCAAAAGACTTTGAGCCAACTTATCAGTATTATATTGAG 547

    115 gLeuPro.FrpLeuAsnArgSerGlu.ThrValValGluGluTyrLeuAl 131
        ||||||| |||||||||||||||||| ||||||||||||||||||||||||
    548 ATTGCCTTTGGTTGAATAGAAGTCAAAACAGTAGTGGAAGAGTATTTGGC 597

    131 aPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysL 148
        |||||||||||||||||||||||||    ||||||||||||||||||||||
    598 TTTTCTTGCTAATCTTGTATCAGCATAGACTGTTTTCCTCAGACCGTGTC 647

    148 euSerMetIleAlaSer.HisPheValProProArgValIleIleLysGl 164
        ||||||||||||||||| ||||||||||||||||||||||||||||||||
    648 TCAGCATGATTGCTTCCCCATTTTGTGCCTCCCCGAGTGATCATTAAGGA 697

    164 uGlyAspValAspValSerAspSerAspAspGluAspAspAsnLeuProA 181
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    698 AGGCGATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCTG 747

    181 laAsnPheAspPhrCysHisArgAlaLeuGlnIleIleAlaArgTyrVal 197
        |||||||||||||||   ||||||||||||||||||||||||||||||||
    748 CAAATTTTGACACATATCACAGAGCCTTGCAAATAATAGCAAGATATGTA 797

    198 ProSerThrProTrpPhe 203
        ||||||      :::|||
    798 CCATCATGAGTATACTTT 815

seq_name: /SIDSI/qcqdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS80955

seq documentation block:
ID   AAS80955 standard; cDNA; 655 BP.
XX
AC   AAS80955;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #16759.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
```

```
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG16768.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 16759; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 655 BP; 207 A; 134 C; 115 G; 199 T; 0 other;


alignment_scores:
            Quality:  831.00              Length:    196
              Ratio:  4.541                Gaps:      6
 Percent Similarity:  93.367   Percent Identity: 92.347

alignment_block:
 US-09-932-678-2 x AAS80955/rev   ..

 Align seq 1/1  to reverse of: AAS80955  from: 1  to: 655

      14 AlaAlaAlaSerSerSerAlaValLysLysLeuGlyAlaSer.ArgThrG 30
         ||||||||||||||||   |||||||||  |||   |||||| |||||||
     655 GCGGCCGCTTCGTCCTGTGCAGTCAAGACGCTGTTCGCGTCGTAGGACTG 606

      30 lyIleSerAsnMetArgAlaLeuGluAsnAspPhePheAsnSerProPro 46
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     605 GGATTTCAAATATGCGTGCATTAGAGAATGATTTTTCAATTCTCCCCCA 556

      47 ArgLysThrValArgPheGlyGlyThrValThrGluValLeuLeuLysTy 63
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     555 AGAAAAACTGTTCCGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTA 506

      63 rLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuA 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     505 CAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAG 456

      80 spProAspIleLysAspAsp.GlnIleIleAsnTrpLeuLeuGluPheAr 96
         |||||||||||||||||| ||||||||||||||||||||||||||||||
     455 ATCCAGACATAAAGGATGACCCAGATCATCAACTGGCTGCTAGAATTCCG 406

      96 qSerSerIleMetTyrLeuThr.LysAspPheGlu.GlnLeuIleSerIl 112
         ||||||||:::|||||||||||| ||||||||||| ||||||||||||||
     405 TTCTTCTGTCATGTACTTGACCAAAAGACTTTGAGCCAACTTATCAGTAT 356

     112 eIleLeuArgLeuProTrpLeuAsnArgSerGln.ThrValValGluGlu 128
         ||||||||||||||||||||||||||||||||| ||||||||||||||||
     355 TATATTGAGATTGCCTTGGTTGAATAGAAGTCAAAACAGTAGTGGAAGAG 306

     129 TyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuAr 145
         ||||||||||||||||||||||||||||||||    ||||||||||||||
     305 TATTTGGCTTTTCTTGGTAATCTTGTATCAGCATAGACTGTTTTCCTCAG 256

     145 qProCysLeuSerMetIleAlaSer.HisPheValProProArgValIle 161
         |||||||||||||||||||||||||| ||||||||||||||||||||||||
     255 ACCGTGTCTCAGCATGATTGCTTCCCCATTTTGTGCCTCCCCGAGTGATC 206

     162 IleLysGluGlyAspValAspValSerAspSerAspAspGluAspAspAs 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     205 ATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGATGATGAAGATGATAA 156

     178 nLeuProAlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaA 195
         ||||||||||||||||||||    ||||||||||||||||||||||||||
     155 TCTTCCTGCAAATTTTGACACATATCACAGAGCCTTGCAAATAATAGCAA 106

     195 rgTyrValProSerThrProTrpPhe 203
         ||||||||||||||||      :::|||
     105 GATATGTACCATCATGAGTATACTTT 80

seq_name: /S1DS1/qcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX51663

seq_documentation_block:
ID   AAX51663 standard; cDNA; 437 BP.
XX
AC   AAX51663;
XX
DT   21-JUN-1999  (first entry)
XX
DE   Human secreted protein 5' EST SEQ ID NO:242.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS   Homo sapiens.
XX
PN   WO9906549-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01231
XX
PR   01-AUG-1997;   97US-0905279.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153779/13.
DR   P-PSDB; AAY12885.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries derived from testis, ovary, uterus and spleen tissue
XX
PS   Claim 1; Page 344-345; 522pp; English.
XX
CC   AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY12681 to
CC   AAY12913, respectively. The proteins given represent the signal peptide
```

```
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ   Sequence 447 BP; 118 A; 85 C; 102 G; 141 T; 1 other;


alignment scores:
            Quality:  712.00              Length:    145
              Ratio:  4.979                 Gaps:      0
 Percent Similarity:  98.621   Percent Identity: 95.862

alignment block:
 US-09-932-678-2 x AAX51663   ..

 Align seq 1/1  to: AAX51663  from: 1  to: 447

      16 AlaSerSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSe 32
         ||||||||||||||||||||||   ||||||||||||||||||||||||||
          3 GCTTCGTCCTCTGCAGTCAAGACGCTGGGCGCGTCGAGGACTGGGATTTC 52

      32 rAsnMetArgAlaLeuGluAsnAspPheAsnSerProProArgLysP 49
         |||||||||||||||||||||||||||||||||||||||||||||||||
      53 AAATATGCGTGCATTAGAGAATGATTTTTTCAATTCTCCCCCAAGAAAAA 102

      49 hrValArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLys 65
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     103 CTGTTCGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAG 152

      66 GlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAs 82
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     153 GGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGA 202

      82 pIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerI 99
         ||||||||||||||||||||||||||||||||||||||||||||||||||:
     203 CATAAAGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTG 252

      99 leMetTyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArg 115
         ::|||||||||||||||||||||||||||||||||||||||||||||||||
     253 TCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTGAGA 302

     116 LeuProTrpLeuAsnArgSerGluThrValValGluGluTyrLeuAlaPh 132
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     303 TTGCCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTT 352

     132 eLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuS 149
         |||||||||||||||||||:::|||||||||| ||||||||||||||||
     353 TCTTGGTAATCTTGTATCAGCAGACGTGTTTCTCAGACCGTGTCTCA 402

     149 erMetIleAlaSerHisPheValProProArgVal 160
         |||||||||||||||||||   ||||||::::::
     403 GCATGATTGCTTCCCATTTTGWGCCTCCCGAGCTG 437

seq name: /SIDSI/seqdata/geneseq/geneseqn_emb1/NA2001B.DAT:AAS91991

seq documentation block:
ID   AAS91991 standard; cDNA; 560 BP.
XX
AC   AAS91991;
XX
DT   18-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #27795.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo Sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-649362/73.
DR   P-PSDB; ABG27804.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   bi-diversity -
XX
PS   Claim 1; SEQ ID No 27795; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 560 BP; 178 A; 96 C; 109 G; 177 T; 0 other;


alignment_scores:
            Quality:  689.00              Length:    163
              Ratio:  4.583                 Gaps:      5
 Percent Similarity:  93.252   Percent Identity: 90.798

alignment block:
 US-09-932-678-2 x AAS91991   ..

 Align seq 1/1  to: AAS91991  from: 1  to: 560

      49 ProArgLysThrValArgPheGlyGlyThrValThrGluValLeuLeuLy 62
         ||||||||||||||::||||||||||||||||||||||||||||||||||
       1 CTAAGAAAAACTGTTCAGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAA 52

      63 sTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuL 79
         |:||||   ||||||||||||||||||||||||||||||||||||||||
       53 GTACAAAACGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGT 102
```

```
     79 euAspProAspIleLysAspAsp.GlnIleIleAsnTrpLeuLeuGluPh 95
        ||||||||   |||:::|||||| ||||||||||||||||||||||||||
    103 TAGATCCAGCCATAAGGGATGACCCAGATCATCAACTGGCTGCTAGAATT 152

     95 eArgSerSerIleMetTyrLeuThr.LysAspPheGlu.GlnLeuIleSe 111
        ||||||||||:::|||||||||||| |||||||||||| ||||||||||
    153 CCGTTCTTCTGTCATGTACTTGACCAAAAGACTTTGAGCCAACTTATCAG 202

    111 rIleIleLeuArgLeuProTrpLeuAsnArgSerGln.ThrValValGlu 127
        |||||||||||||||||||:|||||||||||||||| ||||||||||||
    203 TATTATATTGAGATTGCCTTGGTTGAATAGAAGTCAAAACAGTAGTGGAA 252

    128 GluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLe 144
        ||||||||||||||||||||||||||||||||||||   ||||||||||
    253 GAGTATTTGGCTTTTCTTGGTAATCTTGTATCAGCATAGACTGTTTTCCT 302

    144 uArgProCysLeuSerMetIleAlaSer.HisPheValProProArgVal 160
        |||||||||||||||||||||||||||| ||||||||||||||||||||
    303 CAGACCGTGTCTCAGCATGATTGCTTCCCCATTTTGTGCCTCCCCGAGTG 352

    161 IleIleLysGluGlyAspValAspValSerAspSerAspAspGluAspAs 177
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    353 ATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGATGATGAAGATGA 402

    177 pAsnLeuProAlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleA 194
        ||||||||||||||||||||||   ||||||||||||||||||||||||
    403 TAATCTTCCTGCAAATTTTGACACATATCACAGAGCCTTGCAAATAATAG 452

    194 laArgTyrValProSerThrProTrpPhe 203
        ||||||||||||||||||      ::::||
    453 CAAGATATGTACCATCATGAGTATACTTT 481

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS80954

seq_documentation_block:
ID   AAS80954 standard; cDNA; 396 BP.
XX
AC   AAS80954;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #16758.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG16767.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 16758; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 396 BP; 109 A; 71 C; 96 G; 120 T; 0 other;


alignment_scores:
             Quality:  610.00              Length:     161
               Ratio:  4.841                Gaps:       1
 Percent Similarity:  78.261   Percent Identity: 77.640

alignment_block:
 US-09-932-678-2 x AAS80954   ..

 Align seq 1/1  to: AAS80954  from: 1  to: 396

      17 SerSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSerAs 33
         |||   |||||||||||||   |||||||||       ||||||||||||||
       2 TCGCCCTCTGCAGTCAAGATGCTGGGCGCGTTGAGTACTGGGATTTCAAA 51

      33 nMetArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrV 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      52 TATGCGTGCATTAGAGAATGATTTTTTCAATTCTCCCCCAAGAAAAACTG 101

      50 alArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGly 66
         ||:::|||||||||||||||||||||||||||||||||||||||||   |||
     102 TTCAGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAACGGGT 151

      67 GluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIl 83
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     152 GAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACAT 201

      8- eLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleM 100
         ||||
     2-- AAAG..................................................... 205

     101 etTyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeu 116
                                                    ||||||
     206 .....................................................AGATTG 211

     117 ProTrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLe 133
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     211 CTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCT 261

     133 uGlyAsnLeuValSerAlaGluThrValPheLeuArgProCysLeuSerM 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     262 TGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCA 311

     151 etIleAlaSerHisPheValProProArgValIleIleLysGluGlyAsp 166
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     312 TGATTGCTTCCCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCGAT 361

     167 ValAspValSerAspSerAspAspGluAspAsp 177
         |||||||||||||||||||||||||||||||||
     362 GTAGATGTTTCAGATTCTGATGATGAAGATGAT 394
```

```
OM of: US-09-932-678-2 to: N_Geneseq_032802:*   out_format : pfs

Date: Jul 30, 2002  7:44 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  -MODEL=frame+_p2n.model -DEV=xlh
  -Q=/cgn2_1/USPTO_spool/US09932678/runat_29072002_155156_19107/app_query.fasta_1.717
  -DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=oli.rng -GAPOP=4.500
  -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
  -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
  -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
  -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
  -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
  -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
  -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
  -USER=US09932678_@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
  -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-932-678-2
Query length: 651
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 212.800000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA98384 +   463.00 8547.56  0.0       20
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS44584 +   295.00 5435.62  2.0e-294   1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC76937 +   276.00 5083.83  8.0e-275  14
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS01562 +   240.00 4416.46  1.2e-237   1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAV90248 +    97.00 1776.37  1.3e-90   34
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS87641 +    97.00 1759.62  1.2e-89    3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS73660 +    93.00 1685.90  1.5e-85    2
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS91992 +    93.00 1685.90  1.5e-85    2
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS92253 +    79.00 1436.52  1.1e-71    7
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX51663 +    75.00 1366.69  8.9e-68   43
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS92248 -    75.00 1364.49  1.2e-67    5
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS92254 +    71.00 1277.64  8.1e-63    3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA83038 +    70.00 1264.93  4.1e-62    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS87636 +    69.00 1249.12  3.1e-61    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS92255 +    66.00 1187.02  9.0e-58    2
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS80954 +    63.00 1144.98  2.0e-55    3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS87635 +    63.00 1144.98  2.0e-55    3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS91986 +    63.00 1144.98  2.0e-55    3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS92246 +    63.00 1144.98  2.0e-55    3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK68676 +    62.00 1097.85  8.4e-53    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK68677 +    62.00 1092.92  1.6e-52    3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK85168 +    62.00 1092.92  1.6e-52    3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC77798 +    61.00 1095.36  1.2e-52   21
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS80955 -    59.00 1067.05  4.4e-51    6
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS87643 +    59.00 1062.13  8.2e-51    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK68674 +    58.00 1028.61  6.0e-49    9
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK68678 +    58.00 1021.51  1.5e-48    2
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK85169 -    58.00 1021.51  1.5e-48    2
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS87640 -    54.00  974.69  6.1e-46    6
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS92247 +    45.00  809.59  9.5e-37    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC77533 +    43.00  775.18  7.9e-35   34
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC19091 +    38.00  686.21  7.1e-30   21
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA46215 +    35.00  625.60  1.7e-26    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA56760 +    35.00  625.60  1.7e-26    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA26382 +    35.00  625.60  1.7e-26    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK04888 -    35.00  625.60  1.7e-26    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK30411 +    35.00  625.60  1.7e-26    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAI15019 +    35.00  625.60  1.7e-26    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAI36361 +    35.00  625.60  1.7e-26    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAI04785 +    35.00  625.60  1.7e-26    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS91991 +    32.00  567.72  2.8e-24    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA51325 +    31.00  562.51  5.5e-23    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA69333 +    31.00  562.51  5.5e-23    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA36263 +    31.00  562.51  5.5e-23    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK17615 +    31.00  562.51  5.5e-23    4

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA98384

seq_documentation_block:
ID   AAA98384 standard; cDNA; 2040 BP.
XX
AC   AAA98384;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human RNA polymerase I transcription factor TIF-1A cDNA.
XX
KW   RNA polymerase I transcription factor TIF-1A; antitumor; treatment;
KW   antiproliferative; cell proliferation; cancer; tissue regeneration; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055316-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-DE00767.
XX
PR   17-MAR-1999;   99DE-1011992.
XX
PA   (DEKK-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI   Grummt I, Vingron M;
XX
DR   WPI; 2000-587527/55.
DR   P-PSDB; AAB10936.
XX
PT   New DNA encoding the transcription factor TIF-1A, useful for preventing
PT   or treating diseases associated with abnormal cell proliferation,
PT   particularly tumors -
XX
PS   Claim 1; Fig 2; 38pp; German.
XX
CC   This invention describes a novel DNA sequence (I) that encodes the RNA
CC   polymerase I transcription factor TIF-1A which has antitumor,
CC   antiproliferative and proliferation-inducing activity. The invention also
CC   describes (1) DNA (Ia) encoding a protein (II) with the biological
CC   activity of TIF-1A; (2) a ribozyme (R) corresponding to (I) or (Ia) and
CC   able to bind specifically to, and cleave, its transcribed RNA so as to
CC   reduce or inhibit synthesis of the corresponding protein; (3) an
CC   antisense RNA (AS) with binding properties similar to R; (4) an
CC   expression vector that contains (I), (Ia) or sequences that encode R or
CC   AS; (5) host cells containing the vectors of (4); (6) TIF-1A or (II)
CC   encoded by (I) or (Ia); (7) preparation of TIF-1A or (II) by culturing
CC   cells of (6); (8) ligands that bind to TIF-1A or (II); (9) antagonists
CC   that weaken or block the activity of TIF-1A or (II); (10) a diagnostic
CC   method for detecting abnormal TIF-1A expression; and (11) kit for
CC   carrying out the method in (10). (I), and similar sequences that encode
CC   proteins with equivalent activity, expression vectors containing them, or
CC   their expression products are used to treat or prevent disorders
CC   associated with reduced cellular proliferation, to stimulate cellular
CC   proliferation, and to promote tissue regeneration, e.g. after injury or
CC   radiation therapy. Ribozymes, antisense sequences directed against (I),
CC   also ligands and antagonists of TIF-1A, are used to treat or prevent
CC   disorders associated with excessive cellular proliferation and to inhibit
CC   proliferation, especially in treatment of cancers. (I) and specific
CC   ligands for TIF-1A (particularly antibodies (Ab)) are also useful for
CC   diagnosis of altered TIF-1A expression by (in)direct determination of the
CC   concentration, length and/or sequence of TIF-1A or its mRNA, e.g. for
CC   detecting mutations. Ab can also be used for immunoprecipitation of
CC   TIF-1A and for isolation of related sequences from cDNA expression
```

```
CC   libraries. (i) allows recombinant production of TIF 1A in sufficient
CC   quantities for therapeutic use.
XX
SQ   Sequence 2040 BP; 563 A; 444 C; 474 G; 568 T; 2 other;


alignment scores:
              Quality:  463.00              Length:     652
                Ratio:  0.713                 Gaps:       1
 Percent Similarity:  99.540   Percent Identity: 99.540

alignment block:
 US-09-932-678-2 x AAA98084    ..

 Align seq 1/1   to: AAA98084   from: 1   to: 2040

        1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSe 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       49 ATGGCGGCACCGCTGCTTCACACGCGTTTGCGGGACATGCGGCCGCTTC 89

       17 rSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnM 34
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       90 GTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGACACTGGCATTCAAATA 139

       44 etArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrVal 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      140 TGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTT 189

       51 ArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGlyGl 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      190 CGGTTTGGTGGAACTGTGACAGAAGTCTTGCTCAAGTACAAAAAGGGTGA 239

       67 nThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleL 84
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      240 AACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAA 289

       84 ysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      290 AGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATG 339

      101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuPr 117
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      340 TACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCC 389

      117 GlnLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLeuG 134
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      390 TTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTG 439

      134 lyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMet 150
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      440 GTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATG 489

      151 IleAlaSerHisPheValProProArgValIleIleLysGluGlyAspVa 167
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      490 ATTGCTTCCCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCGATGT 539

      167 lAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPheA 184
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      540 AGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCTGCAAATTTTG 589

      184 spThrCysHisArgAlaLeuGluIleIleAlaArgTyrValProSerThr 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      590 ACACATGTCACAGAGCCTTGCAAATAATAGCAAGATATGTACCATCGACA 649

      201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLy 217
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      650 CCGTGGTTTCTCCATGCCAATACTGGTGGAAAAATTTCCATTTGTTCGAAA 699

      217 sSerGluArgThrLeuGluCysTyrValHisAsnLeu.LeuArgIleSer 233
          ||||||||||||||||||||||||||||||||  ||| ||||||||||||
      700 AFCAGAGACAACACTGGAATGTTACGTTCATNA.GTTACTAAGGATTAGT 748

      234 ValTyrPheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLy 250
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      749 GTATATTTTCCAACCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAA 798

      250 sLeuLeuLysLeuAspValAsnAlaSerArgGluGlyIleGluAspAlaG 267
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      799 ACTACTCAAGTTGGATGTGAATGCATCCCGGCAGGGTATTGAAGATGCTG 848

      267 luGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeu 283
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      849 AAGAAACAGCAACTCAAACTGTGGTGGGACAGATTCCACGGAAGGATTG 898

      284 PheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyPr 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      899 TTTAATATGGATGAAGATGAAGAAACTGAACATGAAACAAAGGCTGGTCC 948

      300 oGluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleL 317
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      949 TGAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCGGCCTGGACATCC 998

      317 euMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspGly 333
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      999 TGATGTCTTTGTTTTGTCCTACATGAAGGATGTCTGCTATGTAGATGGT 1048

      334 LysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnTT 350
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1049 AAGGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACAT 1098

      351 ePheAspLysLeuLeuProThrHisAlaSerCysHisValGlnPheeP 367
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1099 CTTTGACAAACTCCTGTTGCCCACCCATGCCTCCTGCCATGTACAGTTT 1148

      367 heMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeu 383
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1149 TTATGTTTTACCTCTGTAGTTTCAAATTGGGATTCGCAGAGGCATTTTTG 1198

      383 GluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleAr 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1199 GAACATCTCTGGAAAAAATTGCAGGACCCAACTAATCCTGCCATCATCAG 1248

      400 qGluAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheI 417
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1249 GTAGGCTGCTGGAAATATTATTGGAAGCTTTTTGGCAAGAGCTAAATTTA 1298

      417 leProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeu 433
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1299 TTCCTCTTATTACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTG 1348

      434 HisIleTyrLeuAspAsnGlnAspSerGlyThrLysAlaPheCysAspVa 450
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1349 CACATATACCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGT 1398

      450 lAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrP 467
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1399 TGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACNCCT 1448

      467 hrValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeu 483
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1449 TCGTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGTTTG 1498

      484 GlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAs 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1499 CAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAA 1548

      500 nProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleT 517
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1549 TCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTGCTGCAACTCA 1598

      517 hrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsn 533
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1599 CAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAAT 1648

      533 ArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGl 550
          ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
    1639 CGCCAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCA 1688

     550 nIleCysThrAsnProLeuAspThrPhePheProPheAspProCysValL 567
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1689 GATCTGCACAAACCCGCTGGACACCTTCTTCCCCTTTGATCCCTGTGTGC 1738

     567 euLysArgSerLysLysPheIleAspProIleTyrGlnValTrpGluAsp 583
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1739 TGAAGAGGTCAAAGAAATTCATTGATCCTATTTATCAGGTGTGGGAAGAC 1788

     584 MetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIl 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1789 ATGAGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAAGGACAT 1838

     600 eValGluAspGluAspAspAspPheLeuLysGlyGluValProGlnAsnA 617
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1839 AGTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCAGAATG 1888

     617 spThrValIleGlyIleThrProSerSerPheAspThrHisPheArgSer 633
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1889 ATACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAGT 1938

     634 ProSerSerSerValGlySerProProValLeuTyrMetGlnProSerPr 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1939 CCTTCAACTAGTGTGGGCTCCCCACCCGTGTTGTACATGCAACCCAGTCC 1988

     650 oLeu 651
         ||||
    1989 CCTC 1992

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS44584

seq_documentation_block:
ID   AAS44584 standard; DNA; 1498 BP.
XX
AC   AAS44584;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human full-length polynucleotide sequence #9.
XX
KW   Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW   mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW   cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW   nervous system disorder; inflammatory disorder; cell differentiation; ds;
KW   angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW   genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW   cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;
KW   antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW   neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;
KW   immunostimulant; analgesic; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200164834-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04926.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
PR   17-JUN-2000; 2000US-0597707.
PR   14-JUL-2000; 2000US-0616807.
PR   19-SEP-2000; 2000US-0664641.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Wehrman T,  Wang J,  Ma Y,  Wang D,  Chen R,  Xu C;
PI   Drmanac R;
XX
DR   WPI; 2001-589862/66.
DR   P-PSDB; AAU27684.
XX
PT   Novel polypeptides and nucleic acids obtained from cDNA libraries
PT   prepared from various human tissues, for diagnosis, treatment of
PT   cancer, neurological, inflammatory disorders and for use in arrays for
PT   detection -
XX
PS   Claim 1; SEQ ID NO 9; 153pp; English.
XX
CC   Sequences AAS44576-AAS44919 represent full-length polynucleotides and
CC   contig polynucleotides encoding polypeptides of the invention. The DNA
CC   and protein sequences are useful for the treatment, diagnosis and
CC   prevention of various types of disorder in a mammalian subject such as a
CC   human, dog, monkey, mouse, hamster or rat. The disorders include cancers
CC   such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such
CC   as multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC   diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC   disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC   chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC   Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC   disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC   bowel disease. The sequences exhibit activity relating to angiogenesis,
CC   cell proliferation, cell differentiation, stem cell growth factor,
CC   activin or inhibin. Therefore, they can be used to manipulate stem cells
CC   in culture to give rise to neuroepithelial cells that can be used to
CC   augment or replace cells damaged by illness, accidental damage or genetic
CC   disorders. The sequences may also be used for regeneration of bone,
CC   cartilage, tendons and ligaments and in tissue repair and burn healing.
CC   Note: Some sequences for this patent did not form part of the printed
CC   specification, but were obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1498 BP; 417 A; 342 C; 325 G; 424 T; 0 other;


alignment_scores:
            Quality:  295.00              Length:    295
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AAS44584    ..

 Align seq 1/1  to: AAS44584  from: 1  to: 1498

     249 GluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAs 265
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      81 GAAAAACTACTCAAGTTGGATGTGAATGCATCCCGGCAGGGTATTGAAGA 130

     266 pAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluG 282
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     131 TGCTGAAGAAACAGCAACTCAAACTTGTGGTGGGACAGATTCCACGGAAG 180

     282 lyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAla 298
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     181 GATTGTTTAAATATGGATGAAGATGAAGAAACTGAACATGAAACAAAGGCT 230

     299 GlyProGluArgLeuAspGlnMetValHisProValAlaGluArgLeuAs 315
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     231 GGTCCTGAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCGCCTGGA 280

     315 pIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrValA 332
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     281 CATCCTGATGTCTTTGGTTTTGTCCTACATGAAGGATGTCTGCTATGTAG 330

     332 spGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIle 348
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     331 ATGGTAAGGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATA 380

     349 AsnIlePheAspLysLeuLeuProThrHisAlaSerCysHisValGl 365
         |||||||||||||||||||||||||||||||||||||||||||||||||
     381 AACATCTTTGACAAACTCCTGTTGCCCACCCATGCCTCCTGCCATGTACA 430
```

```
      365 nPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaP 482
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      441 GTTTTTCATGTTTTACCTCTGTACTTCAAATTGGGATTGGCAGAGGCAT 490

      482 heLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIle 498
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      491 TTTGGAACATCTCTGGAAAAAATTGCAGGACCCAAGTAATCCTGCCATC 540

      499 IleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLy 415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      541 ATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAGAGCTAA 590

      415 sPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnT 432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      591 ATTTATTCCTCTTATTACTGTAAAATCATGCCTAGATCTTTTGGTTAACT 640

      432 rpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCys 448
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      641 GGCTGCACATATACCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGC 690

      449 AspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTy 465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      691 GATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTA 740

      465 rThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluG 482
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      741 CACTTTGCTTTTTAGACACAAGCAGCTTTGAGCGGAAACCTGAAAGAAG 790

      482 lyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlu 498
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      791 GTTTGCAGTATCTTCAGAGTCTGAATTTGAGCGGATAGTGATGAGCCAG 840

      499 LeuAsnProLeuLysIleCysLeuProSerValValAsnPhePheAlaAl 515
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      841 CTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGC 890

      515 alleThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgA 532
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      891 AATCACAAATAAGTACCAGCTGGTCTTCTGCTACACCATCATCGAGAGGA 940

      532 snAsnArgGlnMetLeuProValIleArgSerThr 563
          |||||||||||||||||||||||||||||||||||
      941 ACAATCGCCAGATGCTGCCAGTCATTAGGAGTACC 975

Seq name: /SIDSI/qcqdata/geneseq/geneseqn_embl/NA2000.DAT:AAC76937

Seq documentation block:
ID   AAC76937 standard; cDNA; 1418 BP.
XX
AC   AAC76937;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF2492 polynucleotide sequence SEQ ID NO:4984.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo Sapiens.
XX
PN   WO200058473-A2.

XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540764.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA, Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB42728.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease
XX
PS   Claim 5; Page 4157-4158; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40247 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 4161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory diseases; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 1418 BP; 422 A; 275 C; 318 G; 403 T; 0 other;


alignment_scores:
            Quality:  276.00              Length:     276
              Ratio:  1.000                 Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AAC76937    ..

 Align seq 1/1  to: AAC76937  from: 1  to: 1418

      275 GlyGlyThrAspSerThrGlnGlyLeuPheAsnMetAspGluAspGluGl 291
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      272 GGTGGGACAGATTCTACGGAAGGATTGTTTAATATGGATGAAGATGAAGA 272

      291 uThrGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetValH 308
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      273 AACTGAACATGAAACAAAGCCTGGTCCTGAACGGCTCGACCAGATGGTGC 322

      308 isProValAlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyr 324
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      323 ATCCTGTAGCCGAGCGCTGGACATCCTGATGTCTTTGGTTTTGTCCTAC 372

      325 MetLysAspValCysTyrValAspGlyLysValAspAsnGlyLysThrLy 341
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      373 ATGAAGGATGTCTGCTATGTAGATGGTAAGGTTGATAACGGCAAAACAAA 422
```

```
     341 sAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProT 358
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     423 GGATCTATATCGCGACCTGATAAACATCTTTGACAAACTCCTGTTGCCCA 472

     358 hrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSerPhe 374
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     473 CCCATGCCTCCTGCCATGTACAGTTTTTCATGTTTTACCTCTGTAGTTTC 522

     375 LysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGl 391
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     523 AAATTGGGATTCGCAGAGGCATTTTTGGAACATCTCTGGAAAAAATTGCA 572

     391 nAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleG 408
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     573 GGACCCAAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATTATATTG 622

     408 lySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSer 424
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     623 GAAGCTTTTTGGCAAGAGCTAAATTTATTCCTCTTATTACTGTAAAATCA 672

     425 CysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAs 441
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     673 TGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCAGGA 722

     441 pSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrS 458
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     723 TTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACT 772

     458 erAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeu 474
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     773 CAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTT 822

     475 LeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPh 491
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     823 TTGAGCGGAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTT 872

     491 eGluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProS 508
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     873 TGAGCGGATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCT 922

     508 erValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPhe 524
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     923 CAGTGGTTAACTTTTTTGCTGCAATCACAAATAAGTACCAGCTCGTCTTC 972

     525 CysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIleAr 541
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     973 TGCTACACCATCATTGAGAGGAACAATCGCCAGATGCTGCCAGTCATTAG 1022

     541 gSerThrAlaGlyGlyAspSerValGln 550
         |||||||||||||||||||||||||||
    1023 GAGTACCGCTGGAGGAGACTCAGTGCAG 1050

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS01562

seq_documentation_block:
ID   AAS01562 standard; cDNA; 1423 BP.
XX
AC   AAS01562;
XX
DT   18-JUL-2001  (first entry)
XX
DE   Human secretory molecule cDNA sptm #52.
XX
KW   Human; secretory molecule; sptm; SPTM; library screening; gene therapy;
KW   cell signalling; cell proliferative disorder; atherosclerosis; cancer;
KW   immune system disorder; AIDS; neurological disorder; Alzheimer's disease;
KW   nervous system disease; mental retardation; developmental disorder;
KW   neuromuscular disorder; microarray; Incyte ID number 4287452dec; ss.
XX
OS   Homo sapiens.
XX
PN   WO200123558-A2.
XX
PD   05-APR-2001.
XX
PF   19-SEP-2000; 2000WO-US25610.
XX
PR   28-SEP-1999;   99US-0156624.
PR   28-SEP-1999;   99US-0156625.
PR   02-DEC-1999;   99US-0168611.
PR   02-DEC-1999;   99US-0168613.
PR   02-DEC-1999;   99US-0168614.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Hodgson DM,  Lincoln SE,  Russo FD,  Spiro PA,  Banville SC;
PI   Bratcher SR,  Dufour GE,  Cohen RJ,  Rosen BH,  Shah P,  Chalup MS;
PI   Hillman JL,  Jones AL,  Yu JY,  Greenawalt LB,  Panzer SR;
PI   Roseberry AM,  Wright RJ,  Chen W,  Liu TF,  Yap PE,  Stockdreher IK;
PI   Amshey S,  Fong WT;
XX
DR   WPI; 2001-258134/25.
XX
PT   New secretory polynucleotides (sptm) and the polypeptides they encode,
PT   for use in inducing antibodies and screening libraries of compounds
XX
PS   Claim 1; Page 152; 161pp; English.
XX
CC   The present sequence for human secretory molecule cDNA sptm #52
CC   (Incyte ID number 4287452dec) is 1 of 63 novel sptm cDNA sequences
CC   (AAS01511-AAS01573) which encode for the secretory polypeptides SPTM. The
CC   sptm polynucleotides are useful for screening a compound for
CC   effectiveness in altering expression of a target polynucleotide, where
CC   the target polynucleotide comprises sptm. Sptm is also useful in a method
CC   for assessing the toxicity of a test compound. Sptm and its fragments or
CC   complementary sequences, may be used to identify the presence of and/or
CC   determine the degree of similarity between two nucleic acid sequences.
CC   Sptm can also be used for a variety of diagnostic and therapeutic
CC   purposes, e.g. diagnosing a particular condition, disease or disorder
CC   associated with cell signalling, such conditions include cell
CC   proliferative disorders such as atherosclerosis, and cancers including
CC   leukaemia, an immune system disorder e.g. acquired immunodeficiency
CC   syndrome (AIDS), a neurological disorder such as epilepsy or Alzheimer's
CC   disease, nutritional and metabolic disease of the nervous system, mental
CC   retardation and other developmental disorders, and muscular dystrophy and
CC   other neuromuscular disorders. Sptm can also be used to design probes
CC   useful in diagnostic assays, which may be used to monitor the progress of
CC   conditions or disorders associated with abnormal levels of expression of
CC   sptm. In addition sptm encoding SPTM may be used for somatic or germline
CC   gene therapy, for inducing antibodies, or in microarrays.
XX
SQ   Sequence 1423 BP; 417 A; 247 C; 320 G; 439 T; 0 other;


alignment_scores:
           Quality:  240.00              Length:    240
             Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AAS01562   ..

 Align seq 1/1  to: AAS01562  from: 1  to: 1423

      31 IleSerAsnMetArgAlaLeuGluAsnAspPhePheAsnSerProProAr 47
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     194 ATTTCAAATATGCGTGCATTAGAGAATGATTTTTTCAATTCTCCCCCAAG 243

      47 gLysThrValArgPheGlyGlyThrValThrGluValLeuLeuLysTyrL 64
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     244 AAAAACTGTTCGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACA 293

      64 ysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     294 AAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGAT 343
```

```
 81 ProAspIleLysAspAspGluIleIleAsnTrpLeuLeuGluPheArgSe 97
    ||||||||||||||||||||||||||||||||||||||||||||||||||
444 CCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTC 493

 97 rSerIleMetTyrLeuThrLysAspPheGluGlnLeuIleSerIleIleL 114
    ||||||||||||||||||||||||||||||||||||||||||||||||||
494 TTCTATCATGTACTTGACAAAAGACTTTGAGCAACTCATCAGTATTATAT 443

114 euArgLeuProTrpLeuAsnArgSerGlnThrValValGluGluTyrLeu 130
    ||||||||||||||||||||||||||||||||||||||||||||||||||
444 TAAGATTGCCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTG 493

131 AlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCy 147
    ||||||||||||||||||||||||||||||||||||||||||||||||||
494 CGTTTCTTGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTG 543

147 sLeuSerMetIleAlaSerHisPheValProProArgValIleIleLysG 164
    ||||||||||||||||||||||||||||||||||||||||||||||||||
544 TCTCAGCATGATTGCTTCCCATTTTGTGCCTCCCCGAGTGATCATTAAGG 593

164 luGlyAspValAspValSerAspSerAspAspGluAspAspAsnLeuPro 180
    ||||||||||||||||||||||||||||||||||||||||||||||||||
594 AAGGCGATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCT 643

181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrVa 197
    ||||||||||||||||||||||||||||||||||||||||||||||||||
644 GCAAATTTGACACATGTCACAGACGCTTGCAAATAATAGCAAGATATGT 693

197 lProSerThrProTrpPheLeuMetProIleLeuValGluLysPheProP 214
    ||||||||||||||||||||||||||||||||||||||||||||||||||
694 ACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCAT 743

214 heValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeu 230
    ||||||||||||||||||||||||||||||||||||||||||||||||||
744 TTGTTCGAAAATCAGAGAGAACACTGCAATGTTACGTTCATAACTTACTA 793

231 ArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLeuIl 247
    ||||||||||||||||||||||||||||||||||||||||||||||||||
794 AGCATTAGTCTATATTTTCCAACCTTGAGGCATGAAATTCTCGAGCTTAT 843

247 eIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleG 264
    ||||||||||||||||||||||||||||||||||||||||||||||||||
844 TATTGAAAAACTACTCAAGTTGGATGTGAATGCATCCCGGCAGGGTATTG 893

264 luAspAlaGluGluThrAla 270
    ||||||||||||||||||
894 AAGATGCTGAACAAACAGCA 913

Seq name: /SIDSI/qcqdata/qeneseq/qeneseqn-embl/NA1999.DAT:AAV90248

Seq documentation block:
ID   AAV90248 standard; cDNA; 441 BP.
XX
AC   AAV90248;
XX
DT   15-FEB-1999  (first entry)
XX
DE   EST clone DI849.
XX
KW   Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW   tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW   receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW   gene therapy; SS.
XX
OS   Homo sapiens.
XX
PN   WO9845456-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06955.
XX
PR   10-APR-1997;   97US-0838821.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  LaVallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070077/06.
XX
PT   New polynucleotides encoding human secreted proteins  derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries.
XX
PS   Claim 1; Page 478; 618pp; English.
XX
CC   The present sequence represents a human expressed sequence tag (EST).
CC   The polynucleotide, which is a secreted EST, and the encoded protein
CC   are predicted to have useful biological activities which would make
CC   them suitable for treating, preventing or ameliorating medical
CC   conditions in humans and animals, although no supporting data is
CC   given. Suggested activities include nutritional activity, immune
CC   stimulating or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The polynucleotide may also be useful for gene therapy.
XX
SQ   Sequence 441 BP; 86 A; 83 C; 80 G; 92 T; 0 other;


alignment_scores:
            Quality:   97.00              Length:      97
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AAV90248    ..

 Align seq 1/1  to: AAV90248  from: 1  to: 441

     474 GlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLe 489
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      22 CAAGTTTTGAAGGGAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCT 71

     489 nAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCysL 506
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      72 GAATTTTGAGCGGATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCC 121

     506 euProSerValValAsnPhePheAlaAlaIleThrAsnLysTyrGluLeu 522
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     122 TGCCCTCAGTGGTTAACTTTTTTGCTGCAATCACAAATAAGTACGAGCTC 171

     523 ValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProVa 539
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     172 GTCTTCTGCTACACCATCATTGAGAGGAACAATCGCCAGATGCTGCCAGT 221

     539 LysArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProL 556
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     222 CATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGC 271

     556 euAspThrPhePheProPheAspProCysValLeuLysArg 569
         |||||||||||||||||||||||||||||||||||||||||
     272 TGGACACCTTCTTCCCCTTTGATCCCTGTGTGCTGAAGAGG 312

Seq name: /SIDSI/qcqdata/qeneseq/qeneseqn-embl/NA2001B.DAT:AAS87641

Seq documentation block:
ID   AAS87641 standard; cDNA; 3169 BP.
XX
AC   AAS87641;
XX
DT   13-FEB-2002  (first entry)
```

```
XX
DE   DNA encoding novel human diagnostic protein #23445.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG23454.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 23445; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3169 BP; 812 A; 841 C; 779 G; 737 T; 0 other;


alignment_scores:
            Quality:   97.00              Length:     97
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AAS87641   ..

 Align seq 1/1  to: AAS87641  from: 1  to: 3169

      386 LeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAl 402
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      857 CTTTGGAAAAAATTGCAGGATCCAAGTAATCCTGCCATCATCAGGCAGGC 906

      402 aAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProL 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      907 TGCTGGAAATTATATTGGAAGCTTTTTGGCAAGAGCTAAATTTATTCCTC 956

      419 euIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIle 435
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      957 TTATTACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATA 1006

      436 TyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLe 452
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1007 TACCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCT 1056

      452 uHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValP 469
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1057 CCATGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTT 1106

      469 heArgHisLysGluLeuLeuSerGlyAsnLeuLysGluGly 482
          ||||||||||||||||||||||||||||||||||||||||||
     1107 TTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGC 1147

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS73660

seq_documentation_block:
ID   AAS73660 standard; cDNA; 2991 BP.
XX
AC   AAS73660;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #9464.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG09473.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 9464; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
```

```
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2991 BP; 655 A; 853 C; 882 G; 601 T; 0 other;


alignment scores:
            Quality:   93.00              Length:      93
              Ratio:   1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-932-678-2 x AAS73660   ..

 Align seq 1/1  to: AAS73660  from: 1  to: 2991

       85 AspAspGluIleIleAsnTrpLeuLeuGluPheArgSerSerIleMetTy 101
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      610 GATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATGTA 659

      101 rLeuThrLysAspPheGluGluLeuIleSerIleIleLeuArgLeuProF 118
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      660 CTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTGAGATTGCCTT 709

      119 rpLeuAsnArgSerGluThrValValGluGluTyrLeuAlaPheLeuGly 134
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      710 GGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTGGT 759

      135 AsnLeuValSerAlaGluThrValPheLeuArgProCysLeuSerMetIl 151
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      760 AATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGAT 809

      151 eAlaSerHisPheValProProArgValIleIleLysGluGlyAspValA 168
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      810 tGCTTCCCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCGATGTAG 859

      168 spValSerAspSerAspAspGluAspAsp 177
          |||||||||||||||||||||||||||||
      860 ATGTTTCAGATTCTGATGATGAAGATGAT 888

seq name: /SIDS1/seqdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS91992

seq documentation block:
ID   AAS91992 standard; cDNA; 2991 BP.
XX
AC   AAS91992;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #27796.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08641.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE ) HYSEQ INC.
XX
PI   Drmanac R1,  Liu C,  Fang YT;
XX
DR   WPI; 2001-649362/73.
DR   P-PSDB; AB027805.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 27796; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2991 BP; 655 A; 853 C; 882 G; 601 T; 0 other;


alignment scores:
            Quality:   93.00              Length:      93
              Ratio:   1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-932-678-2 x AAS91992   ..

 Align seq 1/1  to: AAS91992  from: 1  to: 2991

       85 AspAspGluIleIleAsnTrpLeuLeuGluPheArgSerSerIleMetTy 101
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      610 GATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATGTA 659

      101 rLeuThrLysAspPheGluGluLeuIleSerIleIleLeuArgLeuProI 118
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      660 CTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTGAGATTGCCTT 709

      119 rpLeuAsnArgSerGluThrValValGluGluTyrLeuAlaPheLeuGly 134
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      710 GGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTGGT 759

      135 AsnLeuValSerAlaGluThrValPheLeuArgProCysLeuSerMetIl 151
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      760 AATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGAT 809

      151 eAlaSerHisPheValProProArgValIleIleLysGluGlyAspValA 168
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      810 tGCTTCCCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCGATGTAG 859

      168 spValSerAspSerAspAspGluAspAsp 177
          |||||||||||||||||||||||||||||
      860 ATGTTTCAGATTCTGATGATGAAGATGAT 888

seq name: /SIDS1/seqdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS92253

seq documentation block:
ID   AAS92253 standard; cDNA; 776 BP.
XX
AC   AAS92253;
```

```
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #28057.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG28066.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 1; SEQ ID No 28057; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 776 BP; 206 A; 157 C; 198 G; 215 T; 0 other;


alignment_scores:
            Quality:   79.00              Length:     93
              Ratio:  0.859               Gaps:        1
 Percent Similarity:  98.925   Percent Identity: 98.925

alignment_block:
 US-09-932-678-2 x AAS92253    ..

 Align seq 1/1  to: AAS92253  from: 1  to: 776

      100 MetTyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLe 116
          |||||||||||||||||||||||||||||||||||||||||||||||||||
      499 ATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTGAGATT 548

      116 uPro.TrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPhe 132
          |||| |||||||||||||||||||||||||||||||||||||||||||||
      549 CTTGGTAATCTTGCATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAG 598

      133 LeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSe 149
          ||||||||||||| |||||||||||||||||||||||||||||||||||
      599 CTTGGTAATCTTGCATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAG 648

      149 rMetIleAlaSerHisPheValProProArgValIleIleLysGluGlyA 166
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      649 CATGATTGCTTCCCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCG 698

      166 spValAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsn 182
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      699 ATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCTGCAAAT 748

      183 PheAspThrCysHisArgAlaLeuGln 191
          |||||||||||||||||||||||||||
      749 TTTGACACATGTCACAGAGCCTTGCAA 775

seq_name: /SIDS1/qcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX51663

seq_documentation_block:
ID   AAX51663 standard; cDNA; 437 BP.
XX
AC   AAX51663;
XX
DT   21-JUN-1999  (first entry)
XX
DE   Human secreted protein 5' EST SEQ ID NO:242.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS   Homo sapiens.
XX
PN   WO9906549-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01231.
XX
PR   01-AUG-1997;   97US-0905279.
XX
PA   (GESI ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153779/13.
DR   P-PSDB; AAY12885.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries derived from testis, ovary, uterus and spleen tissue
XX
PS   Claim 1; Page 344-345; 522pp; English.
XX
CC   AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY12681 to
CC   AAY12913, respectively. The proteins given represent the signal peptide
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ ligand activity, anti inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
```

```
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ   Sequence 447 BP; 118 A; 85 C; 102 G; 141 T; 1 other;


alignment scores:
            Quality:   75.00              Length:      75
              Ratio:   1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-932-678-2 x AAX51663    ..

 Align seq 1/1  to: AAX51663  from: 1  to: 447

       24 LeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAs 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
       27 CTGGGCGCGTCGAGGACTGGCATTTCAAATATGCGTGCATTAGAGAATGA 76

       40 pPhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValT 57
          |||||||||||||||||||||||||||||||||||||||||||||||||||
       77 TTTTTTCAATTCTCCCCCAAGAAAAACTGTTCGGTTTGGTGGAACTGTGA 126

       57 hrGluValLeuLeuLysTyrLysLysGlyGluThrAsnAspPheGluLeu 73
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      127 CAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTG 176

       74 LeuLysAsnGlnLeuLeuAspProAspIleLysAspAspGlnIleIleAs 90
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      177 TTGAAGAACCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAA 226

       90 nTrpLeuLeuGlnPheArgSerSer 98
          |||||||||||||||||||||||||
      227 CTGGCTGCTAGAATTCCGTTCTTCT 251

Seq name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS92248

Seq documentation block:
ID   AAS92248 standard; cDNA; 585 BP.
XX
AC   AAS92248;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #28052.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE ) HYSEQ INC
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG28061.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 1; SEQ ID No 28052; 104pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 585 BP; 148 A; 152 C; 123 G; 162 T; 0 other;


alignment scores:
            Quality:   75.00              Length:      75
              Ratio:   1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-932-678-2 x AAS92248/rev    ..

 Align seq 1/1  to reverse of: AAS92248  from: 1  to: 585

       24 LeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAs 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      418 CTGGGCGCGTCGAGGACTGGCATTTCAAATATGCGTGCATTAGAGAATGA 369

       40 pPhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValT 57
          |||||||||||||||||||||||||||||||||||||||||||||||||||
      368 TTTTTTCAATTCTCCCCCAAGAAAAACTGTTCGGTTTGGTGGAACTGTGA 319

       57 hrGluValLeuLeuLysTyrLysLysGlyGluThrAsnAspPheGluLeu 73
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      318 CAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTG 269

       74 LeuLysAsnGlnLeuLeuAspProAspIleLysAspAspGlnIleIleAs 90
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      268 TTGAAGAACCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAA 219

       90 nTrpLeuLeuGlnPheArgSerSer 98
          |||||||||||||||||||||||||
      218 CTGGCTGCTAGAATTCCGTTCTTCT 194

Seq name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS92254

Seq documentation block:
ID   AAS92254 standard; cDNA; 3175 BP.
XX
AC   AAS92254;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #28058.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
```

```
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG28067.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 28058; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3175 BP; 811 A; 849 C; 779 G; 736 T; 0 other;


alignment_scores:
            Quality:   71.00              Length:      71
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AAS92254    ..

 Align seq 1/1  to: AAS92254  from: 1  to: 3175

     200 ThrProFrpPheLeuMetProIleLeuValGluLysProProPheValAr 216
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     491 ACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTCG 540

     216 qLysSerGluArqThrLeuGluCysTyrValHisAsnLeuLeuArqIleS 233
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     541 AAAATCAGAGAGAACACTGGAATGTTACGTTCATAACTTACTAAGGATTA 590

     233 erValTyrProProThrLeuArqHisGluIleLeuGluLeuIleIleGlu 249
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     591 GTGTATATTTTCCAACCTTGAGGCATGAAATTCTGGAGCTTATTATTGAA 640

     250 LysLeuLeuLysLeuAspValAsnAlaSerArqGlnGlyIleGluAspAl 266
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     641 AAACTACTCAAGTTGGATGTGAATGCATCCCGGCAGGGTATTGAAGATGC 690

     266 aGluGluThrAla 270
         ||||||||||||
     691 TGAAGAAACAGCA 703

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA83038

seq_documentation_block:
ID   ABA83038 standard; DNA; 1461 BP.
XX
AC   ABA83038;
XX
DT   05-FEB-2002  (first entry)
XX
DE   Human transcription factor TREX-65 coding sequence.
XX
KW   Human; transcription factor; TREX; cell proliferative disease;
KW   autoimmune disease; inflammation; neurological disease;
KW   developmental disorder; cancer; AIDS; infection; cytostatic; anti HIV;
KW   neuroprotective; antiinflammatory; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200172777-A2.
XX
PD   04-OCT-2001.
XX
PF   13-MAR-2001; 2001WO-US08117.
XX
PR   13-MAR-2000; 2000US-0188986.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Hillman JL,  Baughn MR,  Yue H,  Lal P,  Lu DAM,  Patterson C;
PI   Azimzai Y,  Bandman O,  Tang YT,  Mathur P,  Shah P,  Au-Young J;
PI   Reddy R;
XX
DR   WPI; 2001-570896/64.
DR   P-PSDB; ABB50214.
XX
PT   Novel transcription factor polypeptides, used to treat diseases
PT   associated with altered activity and expression of TREX, and to screen
PT   for agents capable of modulating its activity -
XX
PS   Claim 11; Page 299; 327pp; English.
XX
CC   The present sequence is the coding sequence for a human transcription
CC   factor. The transcription factor and its coding sequence are useful in
CC   the diagnosis, treatment and prevention of diseases associated with
CC   altered expression of the transcription factor e.g. cell proliferative,
CC   autoimmune/inflammatory, neurological and developmental disorders. A
CC   number of specific disorders/diseases are given in the specification,
CC   including: arteriosclerosis, cirrhosis, hepatitis, cancers, AIDS,
CC   allergies, anaemia, asthma, autoimmune thyroiditis, bronchitis, atopic
CC   dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,
CC   Grave's disease, multiple sclerosis, osteoarthritis, pancreatitis,
CC   psoriasis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative
CC   colitis, uveitis, Alzheimer's disease, Huntington's disease, Parkinson's
CC   disease, stroke, and viral, bacterial, fungal and protozoal infections.
XX
SQ   Sequence 1461 BP; 413 A; 335 C; 298 G; 415 T; 0 other;


alignment_scores:
            Quality:   70.00              Length:      70
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x ABA83038    ..

 Align seq 1/1  to: ABA83038  from: 1  to: 1461
```

```
      201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLy 217
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      481 CCCTGGTTTCTCATGCCAATACTGGTGCAAAAATTCCCATTTCTTCGAAA 430

      217 sSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerV 234
          ||||||||||||||||||||||||||||||||||||||||||||||||| |
      431 ATCAGAGAGAACACTGGAATGTTACGTTCATAACTTACTAAGGATTAGTG 380

      234 alTyrPheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLys 250
          |||||||||||||||||||||||||||||||||||||||||||||||||||
      381 TATATTTTCCAACCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAAA 430

      251 LeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGl 267
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      431 CTACTCAAGCTGGATGTCAATGCATCCCGCCAGGGTATTGAAGATGCTGA 480

      267 aGlnThrAla 270
          ||||||||||
      481 AGAAACAGCA 490

Seq name: /SIDS1/gcgdata/geneseq/geneseqn embl/NA2001B.DAT:AAS87636

Seq documentation block:
ID   AAS87636 standard; cDNA; 1017 BP.
XX
AC   AAS87636;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #23440.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08641.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; AB023449.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 1; SEQ ID NO 23440; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197 AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1017 BP; 331 A; 214 C; 230 G; 272 T; 0 other;


alignment scores:
            Quality:   69.00              Length:      69
              Ratio:  1.000               Gaps:         0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09-932-678-2 x AAS87636    ..

 Align seq 1/1  to: AAS87636  from: 1  to: 1017

       40 GlyIleSerAsnMetArgAlaLeuGluAsnAspPheProAsnSerProPr 46
          |||||||||||||||||||||||||||||||||||||||||||||||||||
        7 GGGATTTCAAATATGCGTGCAATAGAGAATGATTTTTTCAATTCTCCCCC 56

       47 nArgLysThrValArgPheGlyGlyThrValThrGluValLeuLeuLysT 63
          |||||||||||||||||||||||||||||||||||||||||||||||||||
       57 AAGAAAAACTGTTCGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGT 106

       64 yrLysLysGlyGluThrAspAspPheGluLeuLeuLysAsnGlnLeuLeu 79
          |||||||||||||||||||||||||||||||||||||||||||||||||||
      107 ACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTA 156

       80 AspProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheAr 96
          |||||||||||||||||||||||||||||||||||||||||||||||||||
      157 GATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAGAATTCCG 206

       96 aSerSer 98
          |||||||
      207 TCCTCT 213

Seq_name: /SIDS1/gcgdata/geneseq/geneseqn embl/NA2001B.DAT:AAS92255

Seq_documentation_block:
ID   AAS92255 standard; cDNA; 2410 BP.
XX
AC   AAS92255;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #28059.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08641.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; AB028068.
```

XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 1; SEQ ID No 28059; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2410 BP; 672 A; 437 C; 547 G; 754 T; 0 other;

```
alignment_scores:
            Quality:   66.00              Length:     66
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AAS92255   ..

 Align seg 1/1  to: AAS92255  from: 1  to: 2410

     427 AspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGl 443
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       1 GATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCAGGATTCGGG 50

     443 yThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaC 460
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      51 AACAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCT 100

     460 ysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSer 476
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGAGC 150

     477 GlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGlu 492
         ||||||||||||||||||||||||||||||||||||||||||||||||
     151 GGAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAG 198
```

```
OM of: US-09-932-678-2 to: Issued_Patents_NA:*   out_format : pts

Date: Jul 30, 2002  7:40 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=x11
 -Q=/cgn2_1/USPTO_spool/US09932678/runat_29072002_155156_19038/app_query.fasta_1.717
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=oli.rni -GAPOP=4.500
 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
 -FRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
 -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
 -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09932678_@CGN1_1_54 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS 1

Search information block:
Query: US-09-932-678-2
Query length: 651
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 52.360000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


score_list:
Sequence              Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-905-223-202 -    8.00  126.58  23.08     309  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-500-569-21 -    8.00  121.60  43.68     603  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-870-518-34 +    8.00  117.17  77.11    1094  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-500-569-9 -    8.00  115.65  93.70    1342  !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-815-809-2 +    8.00  110.07  191.82    2844  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-816-155B-4 +    8.00  110.04  192.59    2856  !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-079-587-4 +    8.00  110.04  192.59    2856  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-326-297-1 +    8.00  107.72  259.30    3901  !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-617-454-1 +    8.00  107.72  259.30    3901  !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US94-01144-1 +    8.00  107.72  259.30    3901
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-815-809-3 -    8.00  103.77  429.94    6628  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-816-155B-5 -    8.00  103.75  431.24    6649  !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-079-587-5 -    8.00  103.75  431.24    6649  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-244-378A-40 -    7.00  122.33  39.81      46  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-049-473-19 -    7.00  122.01  41.46      48  !  S
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-312-648-19 -    7.00  122.01  41.46      48  !  S
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US94-04190-19 -    7.00  122.01  41.46      48
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-400-440A-63 -    7.00  119.20  59.42      70  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-463-093A-63 -    7.00  119.20  59.42      70  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-460-888A-63 -    7.00  119.20  59.42      70  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-894-578-63 -    7.00  119.20  59.42      70  !  S
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-412-017-63 -    7.00  119.20  59.42      70  !  S
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-894-578-187 -    7.00  119.10  60.23      71  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-835-029A-2 +    7.00  116.86  80.31      96  !  S
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-049-511-2 +    7.00  116.86  80.31      96  !  Se
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-934-554-2 +    7.00  116.86  80.31      96  !  Se
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-07877-2 +    7.00  116.86  80.31      96  !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US94-04151-2 +    7.00  116.86  80.31      96  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-426-599B-8 +    7.00  115.38  96.99     117  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-09-032-684-14 -    7.00  107.92  252.49     319  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-225-990-23 +    7.00  107.02  283.36     360  !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-804-439A-1 +    7.00  106.51  302.85     386  !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-804-439A-96 +    7.00  106.51  302.85     386  !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-720-229-1 +    7.00  106.51  302.85     386  !  S
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-720-229-96 +    7.00  106.51  302.85     386  !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-826-532-1 +    7.00  105.74  334.21     428  !  S
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-228-324A-1 +    7.00  105.74  334.21     428  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-905-223-261 +    7.00  105.65  337.93     433  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-067-782A-3 +    7.00  105.55  342.40     439  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-529-878B-38 +    7.00  105.36  350.58     450  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-484-956-70 +    7.00  104.09  412.75     534  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-757-653-70 +    7.00  104.09  412.75     534  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-484-956-69 +    7.00  104.06  414.22     536  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-484-956-71 +    7.00  104.06  414.22     536  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-484-956-72 +    7.00  104.06  414.22     536  !

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-905-223-202

seq_documentation_block:
; Sequence 202, Application US/08905223
; Patent No. 6222029
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Jean-Baptiste D.
;    APPLICANT:  Duelert, Aymeric
;    APPLICANT:  Larroix, Bruno
;    TITLE OF INVENTION:  5' ESTs FOR SECRETED PROTEINS
;    NUMBER OF SEQUENCES:  503
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/905,223
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  202:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  309 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Brain
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  105..252
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 94
;      OTHER INFORMATION:  region 136..283
;      OTHER INFORMATION:  id HSU46355
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  53..83
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 90
;      OTHER INFORMATION:  region 82..112
;      OTHER INFORMATION:  id HSU46355
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  227..276
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 94
```

```
;       OTHER INFORMATION:  region 206..255
;       OTHER INFORMATION:  Id AA011705
;       OTHER INFORMATION:  est
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  109..153
;       IDENTIFICATION METHOD:  Von Heijne matrix
;       OTHER INFORMATION:  score 7.5
;       OTHER INFORMATION:  seq LLLGIALLAYVAS/VW
US-08-905-223-202


alignment scores:
             Quality:     8.00              Length:       8
               Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-932-678-2 x US-08-905-223-202/rev    ..

 Align seq 1/1  to reverse of: US-08-905-223-202  from: 1  to: 309

       14 AlaAlaAlaSerSerSerAlaVal 21
          |||||||||||||||||||||||
      121 GCAGCAGCTTCATCSTCTGCGGTT 98

Seq_name: /eqn2_6/ptodata/1/ina/6B_COMB.seq:US-09-500-569-21

Seq_documentation_block:
; Sequence 21, Application US/09500569
; Patent No. 6329204
; GENERAL INFORMATION:
;  APPLICANT: Calhoon, Rebecca E.
;  APPLICANT: Ratalski, Antoni
;  APPLICANT: Shen, Jennie
;  TITLE OF INVENTION: Plant Caffeic acid 3-O-Methyltransferase Homologs
;  FILE REFERENCE: BB1427 US NA
;  CURRENT APPLICATION NUMBER: US/09/500,569
;  CURRENT FILING DATE: 2000-02-09
;  EARLIER APPLICATION NUMBER: 60/119,587
;  EARLIER FILING DATE: 1999-February-10
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 21
;   LENGTH: 603
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (506)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (508)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (511)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (529)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (531)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (554)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (577)..(578)
US-09-500-569-21


alignment scores:
             Quality:     8.00              Length:       8
               Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-932-678-2 x US-09-500-569-21/rev    ..

 Align seq 1/1  to reverse of: US-09-500-569-21  from: 1  to: 603

       14 AlaAlaAlaSerSerSerAlaVal 21
          |||||||||||||||||||||||
      233 GCGGCGGCTAGCTCGTCCGCGGTC 210

seq_name: /eqn2_6/ptodata/1/ina/5B_COMB.seq:US-08-870-518-34

seq_documentation_block:
; Sequence 34, Application US/08870518
; Patent No. 5925566
;  GENERAL INFORMATION:
;    APPLICANT:  Davis, Roger J.
;    APPLICANT:  Galcheva-Gargova, Zoya
;    TITLE OF INVENTION:  NON ACTIVATED RECEPTOR COMPLEX
;    TITLE OF INVENTION:  PROTEINS AND USES THEREOF
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/870,518
;      FILING DATE:  06-JUN-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/019,219
;      FILING DATE:  06-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, Peter J.
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  04020/102001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1094 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-870-518-34


alignment_scores:
             Quality:     8.00              Length:       8
               Ratio:    1.000               Gaps:        0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-932-678-2 x US-08-870-518-34    ..

 Align seq 1/1  to: US-08-870-518-34  from: 1  to: 1094

      350 IlePheAspLysLeuLeuLeuPro 357
          ||||||||||||||||||||||||
      402 ATATTCGACAAGCTGTTATTGCCA 425

seq_name: /eqn2_6/ptodata/1/ina/6B_COMB.seq:US-09-500-569-9
```

```
seq_documentation_block:
; Sequence 9, Application US/09500569
; Patent No. 6329204
; GENERAL INFORMATION:
;  APPLICANT: Cahoon, Rebecca E.
;  APPLICANT: Kafalski, Antoni
;  APPLICANT: Shen, Jennie
;  TITLE OF INVENTION: Plant Caffeic acid 3-O Methyltransferase Homologs
;  FILE REFERENCE: BB1327 US NA
;  CURRENT APPLICATION NUMBER: US/09/500,569
;  CURRENT FILING DATE: 2000-02-09
;  EARLIER APPLICATION NUMBER: 60/119,587
;  EARLIER FILING DATE: 1999-February-10
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
;   LENGTH: 1342
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-500-569-9


alignment_scores:
            Quality:    8.00            Length:       8
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x US-09-500-569-9/rev   ..

 Align seq 1/1  to reverse of: US-09-500-569-9  from: 1  to: 1342

      14 AlaAlaAlaSerSerSerAlaVal 21
         |||||||||||||||||||||||||
     235 GCGGCGGCCAGCTCGTCCGCGGTC 212

seq_name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-815-809-2

seq_documentation_block:
; Sequence 2, Application US/08815809
; Patent No. 6004777
; GENERAL INFORMATION:
;  APPLICANT: TARTAGLIA, James
;  APPLICANT: GOEBEL, Scott J.
;  APPLICANT: COX, William I.
;  APPLICANT: GETTIG, Russell R.
;  APPLICANT: PINCUS, Steven E.
;  APPLICANT: PAOLETTI, Enzo
;  APPLICANT: JACOBS, Bertram L.
;  TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND METHODS OF
;  TITLE OF INVENTION: MAKING AND USES THEREOF
;  FILE REFERENCE: 454310-3010
;  CURRENT APPLICATION NUMBER: US/08/815,809
;  CURRENT FILING DATE: 1997-03-12
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2844
;   TYPE: DNA
;   ORGANISM: Vaccinia virus
US-08-815-809-2


alignment_scores:
            Quality:    8.00            Length:       8
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x US-08-815-809-2   ..

 Align seq 1/1  to: US-08-815-809-2  from: 1  to: 2844

     541 ArgSerThrAlaGlyGlyAspSer 548
         ||||||||||||||||||||||||
    2584 CGCTCAACGGCTGGTGGGGATAGC 2607

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-816-155B-4

seq_documentation_block:
; Sequence 4, Application US/08816155B
; Patent No. 5990091
;  GENERAL INFORMATION:
;    APPLICANT:  TARTAGLIA, JAMES
;    APPLICANT:  COX, WILLIAM I.
;    APPLICANT:  GETTIG, RUSSELL R.
;    APPLICANT:  MARTINEZ, HECTOR
;    APPLICANT:  PAOLETTI, ENZO
;    APPLICANT:  PINCUS, STEVEN E.
;    TITLE OF INVENTION:  VECTORS HAVING ENHANCED EXPRESSION, AND
;    TITLE OF INVENTION:  METHODS OF MAKING AND USES THEREOF
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;      STREET:  745 FIFTH AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10151
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/816,155B
;      FILING DATE:  12-MAR-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOWALSKI, THOMAS J.
;      REGISTRATION NUMBER:  32,147
;      REFERENCE/DOCKET NUMBER:  454310-2990
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-588-0800
;      TELEFAX:  212-588-0500
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2856 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-816-155B-4


alignment_scores:
            Quality:    8.00            Length:       8
              Ratio:   1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x US-08-816-155B-4   ..

 Align seq 1/1  to: US-08-816-155B-4  from: 1  to: 2856

     541 ArgSerThrAlaGlyGlyAspSer 548
         ||||||||||||||||||||||||
    2590 CGCTCAACGGCTGGTGGGGATAGC 2619

seq_name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-079-587-4

seq_documentation_block:
; Sequence 4, Application US/09079587
; Patent No. 6130066
;  GENERAL INFORMATION:
```

```
;     APPLICANT:  TARTAGLIA, JAMES
;     APPLICANT:  COX, WILLIAM I.
;     APPLICANT:  GETTIG, RUSSELL R.
;     APPLICANT:  MARTINEZ, HECTOR
;     APPLICANT:  PAOLETTI, ENZO
;     APPLICANT:  PINCUS, STEVEN E.
;     TITLE OF INVENTION:  VECTORS HAVING ENHANCED EXPRESSION, AND
;     TITLE OF INVENTION:  METHODS OF MAKING AND USES THEREOF
;     NUMBER OF SEQUENCES:  48
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;       STREET:  745 FIFTH AVENUE
;       CITY:  NEW YORK
;       STATE:  NEW YORK
;       COUNTRY:  USA
;       ZIP:  10151
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/079,587
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/816,155
;       FILING DATE:  12 MAR 1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KOWALSKI, THOMAS J.
;       REGISTRATION NUMBER:  32,147
;       REFERENCE/DOCKET NUMBER:  454410-2990
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-588 0800
;       TELEFAX:  212 588 0500
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2856 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-09-079-587-4


Alignment_scores:
             Quality:    8.00              Length:       8
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-932-678-2 x US-09-079-587-4    ..

 Align_seq 1/1  to: US-09-079-587-4  from: 1  to: 2856

      541 AroSerThrAlaGlyGlyAspSer 548
          ||||||||||||||||||||||||
     2596 CGCTCAACGGCTGGTGGGGATAGC 2619

Seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-326-297-1

Seq_documentation_block:
; Sequence 1, Application US/08326297
; Patent No. 5510253
;  GENERAL INFORMATION:
;     APPLICANT:  Mitsky, Timothy A.
;     APPLICANT:  Tumer, Nilgun E.
;     APPLICANT:  Hemenway, Cynthia L.
;     TITLE OF INVENTION:  Plants Resistant to Infection by PLRV
;     NUMBER OF SEQUENCES:  5
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Janelle D. Strode, Monsanto Co., BB4F
;       STREET:  700 Chesterfield Parkway No. 55102531h
;       CITY:  St. Louis
;       STATE:  Missouri
;       COUNTRY:  USA
;       ZIP:  63198
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/326,297
;       FILING DATE:
;       CLASSIFICATION:  800
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Strode, Janelle D.
;       REGISTRATION NUMBER:  34,738
;       REFERENCE/DOCKET NUMBER:  3B 21(10562)A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (314)537-6224
;       TELEFAX:  (314)537-6047
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8901 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA (genomic)
US-08-326-297-1


Alignment_scores:
             Quality:    8.00              Length:       8
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment_block:
 US-09-932-678-2 x US-08-326-297-1    ..

 Align_seq 1/1  to: US-08-326-297-1  from: 1  to: 8901

      582 GluAspMetSerAlaGluGluLeu 589
          ||||||||||||||||||||||||
     229- GAGGATATGAGCGCAGAAGAGCTG 2316

Seq_name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-617-454-1

Seq_documentation_block:
; Sequence 1, Application US/08617454
; Patent No. 6013864
;  GENERAL INFORMATION:
;     APPLICANT:  Mitsky, Timothy A.
;     APPLICANT:  Tumer, Nilgun E.
;     APPLICANT:  Hemenway, Cynthia L.
;     APPLICANT:  Lawson, E. C.
;     TITLE OF INVENTION:  Plants Resistant to Infection by
;     TITLE OF INVENTION:  Luteovirus
;     NUMBER OF SEQUENCES:  16
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lawrence M. Lavin, Jr., Monsanto Company,
;       ADDRESSEE:  BB4F
;       STREET:  700 Chesterfield Parkway No. 60138641h
;       CITY:  St. Louis
;       STATE:  Missouri
;       COUNTRY:  USA
;       ZIP:  63198
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/617,454
;       FILING DATE:
```

```
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/012,688
;       FILING DATE:  03-FEB-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lavin Jr., Lawrence M.
;       REGISTRATION NUMBER:  30,768
;       REFERENCE/DOCKET NUMBER:  38-21(13583)A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (314)537-6670
;       TELEFAX:  (314)537-6047
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3901 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-617-454-1


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x US-08-617-454-1    ..

 Align seq 1/1  to: US-08-617-454-1  from: 1  to: 3901

     582 GluAspMetSerAlaGluGluLeu 589
         ||||||||||||||||||||||||
    2293 GAGGATATGAGCGCAGAAGAGCTG 2316

seq_name: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US94-01144-1

seq_documentation_block:
; Sequence 1, Application PC/TUS9401144
;  GENERAL INFORMATION:
;     APPLICANT:
;     TITLE OF INVENTION:  Plants Resistant to Infection by PLRV
;     NUMBER OF SEQUENCES:  5
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US94/01144
;       FILING DATE:  01-FEB-1994
;       CLASSIFICATION:
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3901 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA (genomic)
PCT-US94-01144-1


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x PCT-US94-01144-1    ..

 Align seq 1/1  to: PCT-US94-01144-1  from: 1  to: 3901

     582 GluAspMetSerAlaGluGluLeu 589
         ||||||||||||||||||||||||
    2293 GAGGATATGAGCGCAGAAGAGCTG 2316

seq_name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-815-809-3

seq_documentation_block:
; Sequence 3, Application US/08815809
; Patent No. 6004777
; GENERAL INFORMATION:
;  APPLICANT: TARTAGLIA, James
;  APPLICANT: GOEBEL, Scott J.
;  APPLICANT: COX, William I.
;  APPLICANT: GETTIG, Russell R.
;  APPLICANT: PINCUS, Steven E.
;  APPLICANT: PAOLETTI, Enzo
;  APPLICANT: JACOBS, Bertram L.
;  TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND METHODS OF
;  TITLE OF INVENTION: MAKING AND USES THEREOF
;  FILE REFERENCE: 454310-3010
;  CURRENT APPLICATION NUMBER: US/08/815,809
;  CURRENT FILING DATE: 1997-03-12
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 6628
;   TYPE: DNA
;   ORGANISM: Vaccinia virus
US-08-815-809-3


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x US-08-815-809-3/rev    ..

 Align seq 1/1  to reverse of: US-08-815-809-3  from: 1  to: 6628

     541 ArgSerThrAlaGlyGlyAspSer 548
         ||||||||||||||||||||||||
    1251 CGCTCAACGGCTGGTGGGGATAGC 1228

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-816-155B-5

seq_documentation_block:
; Sequence 5, Application US/08816155B
; Patent No. 5990091
;  GENERAL INFORMATION:
;     APPLICANT:  TARTAGLIA, JAMES
;     APPLICANT:  COX, WILLIAM I.
;     APPLICANT:  GETTIG, RUSSELL R.
;     APPLICANT:  MARTINEZ, HECTOR
;     APPLICANT:  PAOLETTI, ENZO
;     APPLICANT:  PINCUS, STEVEN E.
;     TITLE OF INVENTION:  VECTORS HAVING ENHANCED EXPRESSION, AND
;     TITLE OF INVENTION:  METHODS OF MAKING AND USES THEREOF
;     NUMBER OF SEQUENCES:  48
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;       STREET:  745 FIFTH AVENUE
;       CITY:  NEW YORK
;       STATE:  NEW YORK
;       COUNTRY:  USA
;       ZIP:  10151
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/816,155B
```

```
;       FILING DATE:  12 MAR 1997
;       CLASSIFICATION:  514
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KOWALSKI, THOMAS J.
;       REGISTRATION NUMBER:  32,147
;       REFERENCE/DOCKET NUMBER:  454410-2990
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212 588 0800
;       TELEFAX:  212 588-0500
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6649 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-816-155B-5


alignment_scores:
             Quality:    8.00              Length:       8
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x US-08-816-155B-5/rev    ..

 Align seq 1/1  to reverse of: US-08-816-155B-5  from: 1  to: 6649

      541 ArgSerThrAlaGlyGlyAspSer 548
          ||||||||||||||||||||||||
     1272 CGCTCAACGGCTGGTGGGGATAGC 1249

seq_name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-079_587-5

seq_documentation_block:
; Sequence 5, Application US/09079587
; Patent No. 6130066
;  GENERAL INFORMATION:
;     APPLICANT:  TAPIAGLIA, JAMES
;     APPLICANT:  COX, WILLIAM I.
;     APPLICANT:  GETTIG, RUSSELL R.
;     APPLICANT:  MARTINEZ, HECTOR
;     APPLICANT:  PAOLETTI, ENZO
;     APPLICANT:  PINCUS, STEVEN E.
;     TITLE OF INVENTION:  VECTORS HAVING ENHANCED EXPRESSION, AND
;     TITLE OF INVENTION:  METHODS OF MAKING AND USES THEREOF
;     NUMBER OF SEQUENCES:  48
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;       STREET:  745 FIFTH AVENUE
;       CITY:  NEW YORK
;       STATE:  NEW YORK
;       COUNTRY:  USA
;       ZIP:  10151
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/079,587
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/816,155
;       FILING DATE:  12 MAR 1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KOWALSKI, THOMAS J.
;       REGISTRATION NUMBER:  32,147
;       REFERENCE/DOCKET NUMBER:  454410-2990
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-588-0800
;       TELEFAX:  212-588-0500
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6649 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-09-079-587-5


alignment_scores:
             Quality:    8.00              Length:       8
               Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x US-08-244-478A-40/rev    ..

 Align seq 1/1  to reverse of: US-09-079-587-5  from: 1  to: 6649

      541 ArgSerThrAlaGlyGlyAspSer 548
          ||||||||||||||||||||||||
     1272 CGCTCAACGGCTGGTGGGGATAGC 1249

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-244-478A-40

seq_documentation_block:
; Sequence 40, Application US/08244478A
; Patent No. 5756323
;  GENERAL INFORMATION:
;     APPLICANT:  Kallenbach, Sacha
;     APPLICANT:  Doyen, No. 575632Ille
;     APPLICANT:  Rougeon, Francois
;     TITLE OF INVENTION:  METHOD FOR GENERATING STRUCTURAL AND
;     TITLE OF INVENTION:  FUNCTIONAL DIVERSITY IN A PEPTIDE SEQUENCE
;     NUMBER OF SEQUENCES:  62
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
;       STREET:  1755 Jefferson Davis Highway, Fourth Floor
;       CITY:  Arlington
;       STATE:  Virginia
;       COUNTRY:  USA
;       ZIP:  22202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/244,478A
;       FILING DATE:  01-SEP-1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  FR 91/15489
;       FILING DATE:  11 DEC-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Oblon, No. 575632man F.
;       REGISTRATION NUMBER:  24,618
;       REFERENCE/DOCKET NUMBER:  660 078-0 PCT
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  703-413-4000
;       TELEFAX:  703 413-2220
;       TELEX:  248855 OPAT UR
;   INFORMATION FOR SEQ ID NO:  40:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  46 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-244-478A-40
```

```
alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x US-08-244-378A-40/rev    ..

 Align seq 1/1  to reverse of: US-08-244-378A-40  from: 1  to: 46

     541 ArgSerThrAlaGlyGlyAsp 547
         ||||||||||||||||||||
      39 AGGTCGACGGCGGGGGGGAT 19

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-049-473-19

seq_documentation_block:
; Sequence 19, Application US/08049473
; Patent No. 5386021
;  GENERAL INFORMATION:
;    APPLICANT:  Moss, Joel
;    APPLICANT:  Mishima, Koichi
;    APPLICANT:  Nightingale, Maria
;    APPLICANT:  Tsuchiya, Mikako
;    TITLE OF INVENTION:  A MAMMALIAN GUANIN NUCLEOTIDE BINDING
;    TITLE OF INVENTION:  PROTEIN WITH AN ADP-RYBOSYLATION FACTOR DOMAIN
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KNOBBE, MARTENS, OLSON AND BEAR
;      STREET:  620 NEWPORT CENTER DRIVE  SIXTEENTH FLOOR
;      CITY:  NEWPORT BEACH
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/049,473
;      FILING DATE:  19930419
;      CLASSIFICATION:  436
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fuller, Michael L.
;      REGISTRATION NUMBER:  36,516
;      REFERENCE/DOCKET NUMBER:  NIH050.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-235-8550
;      TELEFAX:  619-235-0176
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  48 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-049-473-19


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x US-08-049-473-19/rev    ..

 Align seq 1/1  to reverse of: US-08-049-473-19  from: 1  to: 48

     170 SerAspSerAspAspGluAsp 176
         |||||||||||||||||||||
      39 TCGGACTCGGACGATGAGGAT 19
```

```
OM of: US-09-932-678-2 to: EST:*  out_format : pfs

Date: Jul 30, 2002  6:57 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL-frame+_p2n.model -DEV-xlh
 -Q=/cgn2_1/USPTO_spool/US09932678/runat_29072002_155155_18997/app_query.fasta_1.717
 -DB-EST -QFMT=fastap -SUFFIX=oli.rst -GAPOP=4.500 -GAPEXT=0.050
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15
 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0
 -MAXLEN=2000000000 -USER=US09932678_4CGN1_1_3018 -NCPU=6
 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-932-678-2
Query length: 651
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1679.140000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_est1:AW408066       +   171.00 3107.65  7.5e-164  563  ! AW408066 UI-HF-BM0-adu-b-10-0-U
gb_est2:BG428305       +   169.00 3065.79  1.6e-161 1037  ! BG428305 602498983F1 NIH_MGC_7
gb_est1:AL597171       +   167.00 3034.56  8.9e-160  550  ! AL597171 DKFZp313K0411_r1 313 (
gb_est1:AA213789       +   166.00 3016.47  9.0e-159  535  ! AA213789 zr91a11.s1 NCI_CGAP_GC
gb_est2:BI561523       +   157.00 2848.87  2.0e-149  733  ! BI561523 603256222F1 NIH_MGC_97
gb_est2:BE549643       +   152.00 2761.26  1.5e-144  458  ! BE549643 7b40q08.x1 NCI_CGAP_Lu
gb_est1:AW958173       +   147.00 2667.86  2.4e-139  564  ! AW958173 EST370243 MAGE resequ
gb_est1:AA811628       +   146.00 2651.66  1.9e-138  440  ! AA811628 ob74d03.s2 NCI_CGAP_GC
gb_est2:BG167061       +   141.00 2553.12  5.8e-133  990  ! BG167061 602344740F1 NIH_MGC_89
gb_est1:BE093077       +   140.00 2542.07  2.4e-132  422  ! BE093077 RC5-BT0744-260400-031-
gb_est2:BG611364       +   139.00 2519.47  4.3e-131  697  ! BG611364 602612725F1 NIH_MGC_60
gb_est1:AA481295       -   126.00 2286.44  4.1e-118  380  ! AA481295 zv02e05.r1 NCI_CGAP_GC
gb_est1:AA191111       +   126.00 2285.56  4.6e-118  421  ! AA191111 zp86a03.r1 Stratagene
gb_est2:BE502966       +   125.00 2265.94  5.7e-117  490  ! BE502966 hz81g07.x1 NCI_CGAP_Lu
gb_est2:BF203684       +   122.00 2204.91  1.4e-113  996  ! BF203684 601866331F1 NIH_MGC_17
gb_est1:AW239267       +   120.00 2174.52  7.1e-112  479  ! AW239267 xb38b04.y1 NCI_CGAP_Lu
gb_est1:AV703279       +   114.00 2061.70  1.4e-105  671  ! AV703279 AV703279 ADB Homo sapi
gb_est2:BG149493       +   112.00 2027.59  1.1e-103  498  ! BG149493 nad29e03.x1 NCI_CGAP_L
gb_est2:BG502871       +   109.00 1968.70  2.1e-100  788  ! BG502871 602550430F1 NIH_MGC_61
gb_est2:BE614096       +   106.00 1915.62  1.9e-97   632  ! BE614096 601503834F1 NIH_MGC_71
gb_est1:AW015483       +   103.00 1865.74  1.1e-94   348  ! AW015483 UI-H-BI0-aaq-f-11-0-UI
gb_est1:AW768543       -    93.00 1679.29  2.7e-84   507  ! AW768543 hk62f02.x1 NCI_CGAP_Ly
gb_est2:BG149669       +    92.00 1665.00  1.7e-83   316  ! BG149669 nad32a11.x1 NCI_CGAP_L
gb_est1:AV649508       +    91.00 1645.06  2.2e-82   382  ! AV649508 AV649508 GLC Homo sapi
gb_est2:BI601238       +    88.00 1583.41  6.0e-79   835  ! BI601238 603245755F1 NIH_MGC_96
gb_est2:BI464721       +    86.00 1547.84  5.7e-77   736  ! BI464721 603202341F1 NIH_MGC_97
gb_est2:BG501864       +    80.00 1437.26  8.3e-71   793  ! BG501864 602548963F1 NIH_MGC_61
gb_est1:AW770928       +    75.00 1351.61  4.9e-66   394  ! AW770928 hl93e04.x1 NCI_CGAP_Lu
gb_est1:AI928274       +    75.00 1350.27  5.8e-66   461  ! AI928274 wo67c08.x1 NCI_CGAP_Pr
gb_est2:BF221544       +    75.00 1349.63  6.3e-66   497  ! BF221544 7o69d12.x1 NCI_CGAP_Pr
gb_est2:BF110176       +    75.00 1349.61  6.3e-66   498  ! BF110176 7n50d06.x1 NCI_CGAP_Lu
gb_est1:AW978265       -    74.00 1332.06  6.0e-65   455  ! AW978265 EST390374 MAGE resequ
gb_est2:BF215433       +    73.00 1317.94  3.7e-64   278  ! BF215433 601880963F1 NIH_MGC_57
gb_est2:N42382         +    72.00 1300.26  3.5e-63   258  ! N42382 yy15f11.r1 Soares melanc
gb_est1:AI799954       +    71.00 1276.23  7.7e-62   503  ! AI799954 wc41d02.x1 NCI_CGAP_Pr
gb_est1:AW968124       +    71.00 1275.80  8.1e-62   529  ! AW968124 EST380319 MAGE resequ
gb_est2:BE870692       +    71.00 1272.91  1.2e-61   742  ! BE870692 601448528F1 NIH_MGC_67
gb_est1:AA213873       -    69.00 1240.57  7.5e-60   448  ! AA213873 zr91a11.r1 NCI_CGAP_GC
gb_est2:BI225531       +    68.00 1218.57  1.3e-58   689  ! BI225531 602948775F1 NIH_MGC
gb_est2:BI761160       +    68.00 1217.00  1.5e-58   828  ! BI761160 603043657F1 NIH_MGC
gb_est2:BG545886       +    68.00 1215.87  1.8e-58   946  ! BG545886 602564040F1 NIH_MGC
gb_est2:BE832086       +    66.00 1187.56  6.7e-57   356  ! BE832086 PM0-MT0071-190700-0
gb_est1:AA376316       +    65.00 1172.16  4.8e-56   254  ! AA376316 EST88717 BSC172 cel
gb_est1:AI468181       +    62.00 1118.65  4.6e-54   213  ! AI468181 tq56c02.x1 NCI_CGAP
gb_est2:BI913245       +    62.00 1107.28  2.0e-52   807  ! BI913245 603178986F1 NIH_MGC

seq_name: gb_est1:AW408066

seq_documentation_block:
LOCUS       AW408066                563 bp    mRNA    linear   EST 16-FEB-2000
DEFINITION  UI-HF-BM0-adu-b-10-0-UI.r1 NIH_MGC_38 Homo sapiens cDNA clone
            IMAGE:3062754 5', mRNA sequence.
ACCESSION   AW408066
VERSION     AW408066.1  GI:6927123
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 563)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Eco RI site shown at the beginning of the sequence.
             Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: M.B. Soares Lab
             cDNA Library Arrayed by: M.B. Soares Lab
             DNA Sequencing by: M.B. Soares Lab
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1..563
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3062754"
                     /clone_lib="NIH_MGC_38"
                     /tissue_type="lymph"
                     /cell_type="germinal center B cells"
                     /cell_line="MGC85"
                     /lab_host="DH10B (LTI)"
                     /note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI:
                     Constructed from size fractionated cytoplasmic mRNA
                     (2.5-3.5kb). Directionally cloned. Cells provided by Louis
                     M. Staudt, Ph.D. Library preparation by Maria de Fatima
                     Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT      151 a    122 c    119 g    171 t
ORIGIN


alignment_scores:
           Quality:  171.00              Length:    185
             Ratio:  0.929                 Gaps:      1
 Percent Similarity:  99.459   Percent Identity: 99.459

alignment_block:
 US-09-932-678-2 x AW408066  ..

 Align seq 1/1  to: AW408066  from: 1  to: 563

     330 TyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAs 346
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      10 TATGTAGATGGTAAGGTTGATAACGGCAAAACAAAGGATCTATATCGCGA 59

     347 pLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysH 393
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      60 CCTGATAAACATCTTTGACAAACTCCTGTTGCCCACCCATGCCTCCTGCC 109
```

```
      363 tsValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAla 379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      110 ATGTACAGTTTTTCATGTTTACCTCTGTAGTTTCAAATTGGGATTCGCA 159

      380 GluAlaPheLeuGlnHisLeuTrpLysLysLeuGlnAspProSerAsnPr 396
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      160 GAGGCATTTTTGCAACATCTCTGGAAAAAATTGCAGGACCCAAGTAATCC 209

      396 oAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaA 413
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      210 TGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAA 259

      413 rgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeu 429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      260 CAGCTAAATTTATTCCTCTTATTACTGTAAAATCATGCCTAGATCTTTTG 309

      430 ValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAl 446
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      310 GTTAACTGGCTGCACATATACCTTAATAACCAGGATTCGGGAACAAAGGC 359

      446 aPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaV 463
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      360 ATTCTGCGATCTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTG 409

      463 alPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeu 479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      410 TGTTCTACACCTTTCTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTG 459

      480 LysGluGlyLeuGln.TyrLeuGlnSerLeuAsnPheGluArgIleValM 496
          |||||||||||||||| ||||||||||||||||||||||||||||||||||
      460 AAAGAAGGTTTGCAGGTACCTTCAGACTCTGAATTTTCAGCGGATAGTGA 509

      496 etSerGluLeuAsnProLeuLysIle'ysLeuProSerValValAsnPhe 512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      510 TCAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTT 559

      513 Phe 515
          |||
      560 TTT 562

seq_name: gb_est2:BG428305

seq_documentation_block:
LOCUS       BG428305                1037 bp    mRNA    linear   EST 14-MAR-2001
DEFINITION  602498984F1 NIH MGC 75 Homo sapiens cDNA clone IMAGE:4612605 5',
            mRNA sequence.
ACCESSION   BG428305
VERSION     BG428305.1  GI:13384811
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1037)
  AUTHORS   NIH_MGC_http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            cDNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1361  row: e  column: 22
            High quality sequence stop: 647.
FEATURES             Location/Qualifiers
     source          1..1037
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4612605"
                     /clone_lib "NIH MGC 75"
                     /lab_host "DH10B (T1 phage resistant)"
                     /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgactctagcc); Site_2: SfiI (gaccattatgccc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(40)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T).  Average insert size 1.65
                     kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
                     by PCR. This library was enriched for full length clones
                     and was constructed by Clontech Laboratories (Palo Alto,
                     CA). Note: this is a NIH MGC Library."
BASE COUNT       292 a    218 c    264 g    264 t
ORIGIN


alignment_scores:
             Quality:  169.00              Length:    184
               Ratio:  0.929                 Gaps:      1
 Percent Similarity:  99.454   Percent Identity: 99.454

alignment_block:
 DS-09-932-678-2 x BG428305   ..

 Align seq 1/1  to: BG428305  from: 1  to: 1037

       1 M etAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSe 17
         |.|||||||||||||||||||||||||||||||||||||||||||||||||
      22 ATGGCGGCACCGCTCCTTCACACGCGTTTGCCCGGCAGATGCGGCCGCTTC 71

      17 rSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnM 34
         |r||||||||'|||||||||||||||||||||||||||||||||||||||||
      72 GTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATA 121

      34 etArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrVal 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     122 TACGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCCAAGAAAAACTGTT 171

      51 ArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGlyGl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     172 CGGTTTGGTGGAACTGTGACAGAAGTCTTGCTCAAGTACAAAAAGGGTGA 221

      67 nThrAsnAspPheGluLeuLeuLysAsnGluLeuLeuAspProAspIleL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     222 AACAAATGACTTTGAGTTGTTGAAGAACGAGCTGTTAGATCCAGACATAA 271

      84 yrAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     272 TAGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATG 321

     101 TyrLeuThrLysAspPheGluGluLeuIleSerIleIleLeuArgLeuPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     322 TACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCC 371

     117 eTrpLeuAsnArgSerGluThrValValGlnGluTyrLeuAlaPheLeuG 134
         |r||||||||||||||||||||||||||||||||||||||||||||||||||
     372 TTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTG 421

     134 lyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMet 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     422 GTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATG 471

     151 IleAlaSerHisPheValProProArg.ValIleIleLysGluGlyAspV 167
         ||||||||||||||||||||||||||| ||||||||||||||||||||||||
     472 ATTGCTTCCCATTTTGTGCCTCCCCGAAGTGATCATTAAGGAAGGCGATG 521

     167 alAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsn 182
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     522 TAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCTGCAAAT 568

seq_name: gb_est1:AL597171

seq_documentation_block:
```

```
LOCUS       AL597171                 550 bp    mRNA    linear   EST 14-AUG-2001
DEFINITION  DKFZp313K0411_r1 313 (synonym: hlcc2) Homo sapiens cDNA clone
            DKFZp313K0411 5', mRNA sequence.
ACCESSION   AL597171
VERSION     AL597171.1  GI:15154983
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Koehrer,K., Beyer,A., Mewes,W., Weil,B. and Wiemann,S.
  TITLE     EST (Koehrer,K., Beyer,A., Mewes,H.W., Weil,B. and Wiemann,S.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Koehrer K
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert.
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by BMFZ (Biomedical Research Center at the Charite,
            Berlin/Germany) within the cDNA sequencing consortium of the German
            Genome Project.
            No sl sequence available.
            This clone (DKFZp313K0411) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1..550
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp313K0411"
                     /clone_lib="313 (synonym: hlcc2)"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
BASE COUNT      140 a    116 c    140 g    153 t      1 others
ORIGIN


alignment_scores:
            Quality:  167.00            Length:    167
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AL597171   ..

 Align seq 1/1  to: AL597171  from: 1  to: 550

       1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      48 ATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGATGCCGCCGCTTC 97

      17 rSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnM 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      98 GTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATA 147

      34 etArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrVal 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     148 TGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTT 197

      51 ArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGlyGl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     198 CGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGA 247

      67 uThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     248 AACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAA 297

      84 ysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     298 AGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATG 347

     101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     348 TACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCC 397

     117 oTrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLeuG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     398 TTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTG 447

     134 lyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMet 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     448 GTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATG 497

     151 IleAlaSerHisPheValProProArgValIleIleLysGluGlyAspVa 167
         |||||||||||||||||||||||||||||||||||||||||||||||||
     498 ATTGCTTCCCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCGATGT 547

     167 l 167
         |
     548 A 548

seq_name: db_est1:AA213789

seq_documentation_block:
LOCUS       AA213789                 535 bp    mRNA    linear   EST 13-AUG-1997
DEFINITION  zr91a11.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:683036 3',
            mRNA sequence.
ACCESSION   AA213789
VERSION     AA213789.1  GI:1812416
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 535)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: polyT not found
            Insert Length: 1076  Std Error: 0.00
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 457.
FEATURES             Location/Qualifiers
     source          1..535
                     /organism="Homo sapiens"
                     /db_xref="GDB:5586356"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:683036"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTT 3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      150 a    102 c    121 g    161 t      1 others
ORIGIN
```

```
alignment_scores:
            Quality:  166.00              Length:     166
              Ratio:  1.000                 Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-932-678-2 x AA214789    ..

 Align seq 1/1  to: AA214789  from: 1  to: 535

       16 AlaSerSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSe 32
          ||||||||||||||||||||||||||||||||:|||||||||||||||||
           8 GCTTCGTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTC 57

       42 rAsnMetArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysT 49
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       58 AAATATGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAA 107

       49 hrValArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLys 65
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      108 CTGTTCGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAG 157

       66 GlyGluThrAsnAspPheGluLeuLeuLysAsnGluLeuLeuAspProAs 82
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      158 GGTGAAACAAATGACTTTGAGTTGTTGAAGAACGAGCTGTTAGATCCAGA 207

       82 pIleLysAspAspGluIleIleAsnTrpLeuLeuGluPheArgSerSerI 99
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      208 CATAAAGGATGACGAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTA 257

       99 leMetTyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArg 115
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      258 TCATGTACTTCACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGA 307

      116 LeuProTrpLeuAsnArgSerGluThrValValGluGluTyrLeuAlaPh 132
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      308 TTGCCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTT 457

      132 eLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuS 149
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      358 TCTTGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCA 407

      149 erMetIleAlaSerHisPheValProProArgValIleIleLysGluGly 165
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      408 GCATGATTGCTTCCCATTTTGTGCCTCCCCGAGTCATCATTAAGGAAGGC 457

      166 AspValAspValSerAspSerAspAspGluAspAspAsnLeuProAla 181
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      458 GATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCTGCA 505

seq name: qb_est2:BI561523

seq documentation block:
LOCUS       BI561523                 733 bp    mRNA    linear   EST 05-SEP 2001
DEFINITION  603256222F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5298581 5',
            mRNA sequence.
ACCESSION   BI561523
VERSION     BI561523.1  GI:15448847
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 733)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11755  row: I  column: 06
            High quality sequence stop: 717.
FEATURES             Location/Qualifiers
     source          1..733
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5298581"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI XhoI (GtcGac
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN 3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH MGC Library."
BASE COUNT       205 a    142 c    171 g    214 t      1 others
ORIGIN


alignment_scores:
            Quality:  157.00              Length:     157
              Ratio:  1.000                 Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment block:
 US-09-932-678-2 x BI561523    ..

 Align seq 1/1  to: BI561523  from: 1  to: 733

        1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSe 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       23 ATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGATGCGGCCGCTTC 72

       17 rSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnM 34
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       73 GTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATA 122

       34 eLArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrVal 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      123 TGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTT 172

       51 A.qPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGlyGl 67
          |:||||||||||||||||||||||||||||||||||||||||||||||||
      173 CGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGA 222

       67 nThrAsnAspPheGluLeuLeuLysAsnGluLeuLeuAspProAspIleL 84
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      223 AACAAATGACTTTGAGTTGTTGAAGAACGAGCTGTTAGATCCAGACATAA 272

       84 ysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      273 AGGATGACCAGATCATCAACTGGTTGCTAGAATTCCGTTCTTCTATCATG 322

      101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuPr 117
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      323 TACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCC 372

      117 eTrpLeuAsnArgSerGluThrValValGluGluTyrLeuAlaPheLeuG 134
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      373 TTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTG 422

      134 lyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMet 150
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      423 GTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATG 472
```

```
     151 IleAlaSerHisPheValPro 157
         |||||||||||||||||||||
     473 ATTGCTTCCCATTTTGTGCCT 493

seq_name: qb_est2:BE549643

seq_documentation_block:
LOCUS       BE549643                458 bp    mRNA    linear   EST 10-AUG-2000
DEFINITION  7b40g08.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3230750 3'
            similar to TR:O75703 O75703 HYPOTHETICAL 12.4 KD PROTEIN. ;, mRNA
            sequence.
ACCESSION   BE549643
VERSION     BE549643.1  GI:9791335
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 458)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .458
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3230750"
                     /clone_lib="NCI_CGAP_Lu24"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Plasmid DNA from the normalized
                     library NCI_CGAP_Lu5 was prepared, and ss circles were
                     made in vitro. Following HAP purification, this DNA was
                     used as tracer in a subtractive hybridization reaction.
                     The driver was PCR-amplified cDNAs from a pool of 5,000
                     clones made from the same library (cloneIDs
                     1414920-1417991 and 1520904-1522439). Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT      133 a     82 c    100 g    143 t
ORIGIN


alignment_scores:
            Quality:  152.00             Length:    152
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x BE549643   ..

 Align seq 1/1  to: BE549643  from: 1  to: 458

      30 GlyIleSerAsnMetArgAlaLeuGluAsnAspPhePheAsnSerProPr 46
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       3 GGGATTTCAAATATGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCC 52

      46 oArgLysThrValArgPheGlyGlyThrValThrGluValLeuLeuLysT 63
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      53 AAGAAAAACTGTTCGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGT 102

      63 yrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeu 79
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     103 ACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTA 152

      80 AspProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheAr 96
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     153 GATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAGAATTCCG 202

      96 qSerSerIleMetTyrLeuThrLysAspPheGluGlnLeuIleSerIleI 113
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     203 TTCTTCTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTA 252

     113 leLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGluTyr 129
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     253 TATTAAGATTGCCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTAT 302

     130 LeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgPr 146
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     303 TTGGCTTTTCTTGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACC 352

     146 oCysLeuSerMetIleAlaSerHisPheValProProArgValIleIleL 163
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     353 GTGTCTCAGCATGATTGCTTCCATTTTGTGCCTCCCCGAGTGATCATTA 402

     163 ysGluGlyAspValAspValSerAspSerAspAspGluAspAspAsnLeu 179
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     403 AGGAAGGCGATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTT 452

     180 ProAla 181
         ||||||
     453 CCTGCA 458

seq_name: qb_est1:AW958173

seq_documentation_block:
LOCUS       AW958173                564 bp    mRNA    linear   EST 01-JUN 2000
DEFINITION  EST370243 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW958173
VERSION     AW958173.1  GI:8147856
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 117
            Seq primer: Reverse.
FEATURES             Location/Qualifiers
     source          1. .564
                     /organism "Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGE"
                     /note="Vector: pBluescriptSKm"
BASE COUNT      157 a    108 c    130 g    168 t      1 others
ORIGIN


alignment_scores:
            Quality:  147.00             Length:    147
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment block:
 US-09-932-678-2 x AW958173  ..

 Align Seq 1/1  to: AW958173  from: 1  to: 564

       6 LeuHisThrArgLeuProGlyAspAlaAlaAlaSerSerSerAlaValLy 22
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       3 CTTCACACGCGTTTGCCGGGAGATGCGGCCGCTTCGTCCTCTGCAGTTAA 52

      23 sLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluA 39
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      53 GAAGCTGGGCCGCGAGGACTGGGATTTCAAATATGCGTGCCATTAGAGA 102

      40 snAspPhePheAsnSerProProArgLysThrValArgPheGlyGlyThr 55
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     103 ATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTTCGGTTTGGTGGAACT 152

      56 ValThrGluValLeuLeuLysTyrLysLysGlyGluThrAsnAspPheGl 72
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     153 GTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGA 202

      72 uLeuLeuLysAsnGlnLeuLeuAspProAspIleLysAspAspGlnIleI 89
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     203 GTTGCTGAAGAACCAGCTGCTAGATCCAGACATAAAGGATGACCAGATCA 252

      89 leAsnTrpLeuLeuGluPheArgSerSerIleMetTyrLeuThrLysAsp 105
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     253 TCAACTGGCTGCTAGAATTCCGTTCTTCTATCATGTACTTGACAAAAGAC 302

     106 PheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgSe 122
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     303 TTTGAGCAACTTATCAGTATTATATATTAAGATTGCCTTGGTTGAATAGAAG 352

     122 rGluThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerA 139
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     353 TCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTGCTAATCTTGTATCAG 402

     139 laGlnThrValPheLeuArgProCysLeuSerMetIleAla 152
         ||||||||||||||||||||||||||||||||||||||||
     403 CACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCT 444

seq_name: gb_est1:AA811628

seq_documentation block:
LOCUS       AA811628                440 bp    mRNA    linear   EST 19-FEB-1998
DEFINITION  ob74d03.s2 NCI_CGAP_GCB1 Homo Sapiens cDNA clone IMAGE:1337093 3',
            mRNA sequence.
ACCESSION   AA811628
VERSION     AA811628.1  GI:2881239
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo Sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 440)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
            Insert Length: 824  Std Error: 0.00
            Seq primer:  40mld fwd. EF from Amersham
            High quality sequence stop: 424.
FEATURES             Location/Qualifiers
     source          1..440
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1337093"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTT 3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT       130 a      82 c      95 g     133 t
ORIGIN


alignment_scores:
             Quality:  146.00              Length:    146
               Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AA811628    .

 Align Seq 1/1  to: AA811628  from: 1  to: 440

       4 AsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluVa 59
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       1 AATTCTCCCCCTAAGAAAAACTGTTCGGTTTGGTGGAACTGTGACAGAAGT 50

      59 lLeuLeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysA 76
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      51 CTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGA 100

      77 snGlnLeuLeuAspProAspIleLysAspAspGlnIleIleAsnTrpLeu 92
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     101 ACAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTG 150

      93 LeuGluPheArgSerSerIleMetTyrLeuThrLysAspPheGluGlnLe 109
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     151 CTAGAATTCCGTTCTTCTATCATGTACTTGACAAAAGACTTTGAGCAACT 200

     109 uIleSerIleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrValV 126
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     201 TATCAGTATTATATATTAAGATTGCCTTGGTTGAATAGAAGTCAAACAGTAG 250

     127 alGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrVal 142
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     251 TCGAAGAGTATTTGGCTTTTCTTGCTAATCTTGTATCAGCACAGACTGTT 300

     143 PheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProAr 159
         ||||||||:|||||||||||||||||||||||||||||||||||||||||||
     301 TTCCTCAGACCGTCTCTCAGCATGATTGCTTCCCATTTTGTGCCTCCCCG 350

     159 gValIleIleLysGluGlyAspValAspValSerAspSerAspAspGluA 176
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     351 ATCATCATTAAGGAAGGGGATGTAGATGTCTCAGATTCTGATGATGAAG 400

     176 snAspAsnLeuProAlaAsnPheAspThrCysHisArg 188
         ||||||||||||||||||||||||||||||||||||||
     401 AGATAATCTTCCTGCAAATTTTGACACATGTCACAGA 438

seq_name: gb_est2:BG167061
```

```
seq_documentation_block:
LOCUS       BG167061                990 bp    mRNA    linear   EST 06-FEB-2001
DEFINITION  602344740F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4455009 5',
            mRNA sequence.
ACCESSION   BG167061
VERSION     BG167061.1  GI:12673764
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 990)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10247  row: o  column: 10
            High quality sequence stop: 665.
FEATURES             Location/Qualifiers
     source          1. .990
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4455009"
                     /clone_lib="NIH_MGC_89"
                     /tissue_type="hypernephroma, cell line"
                     /lab_host "DH10B (phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT       304 a    275 c    219 g    192 t
ORIGIN


alignment_scores:
            Quality:  141.00             Length:    141
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x BG167061   ..

 Align seq 1/1  to: BG167061  from: 1  to: 990

     501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleTh 517
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     108 CCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGCAATCAC 157

     517 rAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnA 534
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     158 AAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAATC 207

     534 rqGlnMetLeuProValIleArqSerThrAlaGlyGlyAspSerValGln 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     208 GCCAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAG 257

     551 IleCysThrAsnProLeuAspThrPhePheProPheAspProCysValLe 567
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     258 ATCTGCACAAACCCGCTGGACACCTTCTTCCCCTTTGATCCCTGTGTGCT 307

     567 uLysArqSerLysLysPheIleAspProIleTyrGlnValTrpGluAspM 584
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     308 GAAGAGGTCAAAGAAATTCATTGATCCTATTTATCAGGTGTGGGAAGACA 357

     584 etSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     358 TGAGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAAGGACATA 407

     601 ValGluAspGluAspAspAspPheLeuLysGlyGluValProGlnAsnAs 617
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     408 GTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCAGAATGA 457

     617 pThrValIleGlyIleThrProSerSerPheAspThrHisPheArqSerP 634
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     458 TACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAGTC 507

     634 roSerSerSerValGlySerPro 641
         |||||||||||||||||||||||
     508 CTTCAAGTAGTGTGGGCTCCCCA 530

seq_name: gb_est1:BE093077

seq_documentation_block:
LOCUS       BE093077                422 bp    mRNA    linear   EST 12 JUN 2000
DEFINITION  RC5-BT0744-260400-031 H10 BT0744 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE093077
VERSION     BE093077.1  GI:8483529
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 422)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?l1 &t2=RC5 BT0744-260
            400-031-H1&st3-2000-04-26&t4-1)
            Seq primer: puc 18 forward
            High quality sequence stop: 414.
FEATURES             Location/Qualifiers
     source          1. .422
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0744"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       121 a     98 c     99 g    104 t
ORIGIN


alignment_scores:
            Quality:  140.00             Length:    140
```

```
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-932-678-2 x BE093077   ..

 Align seq 1/1   to: BE093077   from: 1   to: 422

      499 LeuAsnProLeuLysIleCysLeuProSerValValAsnPhePheAlaAl 515
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
        3 CTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGC 52

      515 alleThrAsnLysTyrGluLeuValPheCysTyrThrIleIleGluArgA 532
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
       54 AATCACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGA 102

      533 snAsnArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSer 548
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      104 ACAATCGCCAGATGCTGCCAGTCATTACGAGTACCGCTGGAGGAGACTCA 152

      549 ValGlnIleCysThrAsnProLeuAspThrPhePheProPheAspProCy 565
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      153 GTGCAGATCTGCACAAACCCGCTGGACACCTTCTTCCCCTTCGATCCCTG 202

      565 sValLeuLysArgSerLysLysPheIleAspProIleTyrGlnValTrpG 582
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      203 TGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTTATCAGGTGTGGG 252

      582 luAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLys 598
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      253 AACACATGAGTGCTGAACAGCTACAGGAGTTCAAGAAACCCATGAAAAAG 302

      599 AspIleValGluAspGluAspAspAspPheLeuLysGlyGluValProGl 615
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      303 GATATAGTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCA 352

      615 nAsnAspThrValIleGlyIleThrProSerSerPheAspThrHisPheA 632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      353 GAATGATACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTTCC 402

      632 rgSerProSerSerSerVal 638
          ||||||||||||||||||||
      403 GAAGTCCTTCAATAGTGTG 422

seq name: gb_est2:BG611364

seq documentation block:
LOCUS       BG611364                 697 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602612725F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4738063 5',
            mRNA sequence.
ACCESSION   BG611364
VERSION     BG611364.1  GI:13662735
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 697)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1607  row: a  column: 08
            High quality sequence stop: 667.
FEATURES             Location/Qualifiers
     source          1..697
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4738063"
                     /clone_lib="NIH_MGC_60"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (T1 phage resistant)"
                     /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgaccteggcc); Site_2: SfiI (ggccattatggcc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dI(30)BN-3'
                     (where B = G, C, or G and N = A, C, G, or T).  Average
                     insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH MGC
                     Library."
BASE COUNT       178 a    170 c    170 g    179 t
ORIGIN


Alignment scores:
            Quality:  139.00              Length:     139
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

Alignment block:
 US-09-932-678-2 x BG611364   ..

 Align seq 1/1   to: BG611364   from: 1   to: 697

      515 ProAlaIleIleThrAsnLysTyrGluLeuValPheCysTyrThrIleIl 529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
       81 TTGCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCAT 130

      529 eGluArgAsnAsnArgGluMetLeuProValIleArgSerThrAlaGlyG 546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      131 TGACAGGAACAATCGCCAGATGCTGCCAGTCATTACGAGTACCGCTGGAG 180

      546 lyAspSerValGlnIleCysThrAsnProLeuAspThrPhePheProPhe 562
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      181 GAGACTCAGTGCACATCTGCACAAACCCGCTGGACACCTTCTTCCCCTTT 230

      562 AspProCysValLeuLysArgSerLysLysPheIleAspProIleTyrGl 579
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      231 GATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTTATCA 280

      579 nValTrpGluAspMetSerAlaGluGluLeuGlnGluPheLysLysProM 596
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      281 GGTGTGGGAAGACATGAGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCA 330

      596 etLysLysAspIleValGluAspGluAspAspAspPheLeuLysGlyGlu 612
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      331 TGAAAAAGGACATAGTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAA 380

      613 ValProGlnAsnAspThrValIleGlyIleThrProSerSerPheAspTh 629
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      381 GTGCCCCAGAATGATACCGTGATTGGGATCACACCAAGCTCCTTTGACAC 430

      629 rHisPheArgSerProSerSerSerValGlySerProProValLeuTyrM 646
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      431 CATTTCCGAAGTCCTTCAATAGTGTGGGCTCCCCACCCGTGTTGTACA 480

      646 etGlnProSerProLeu 651
          |||||||||||||||
      481 TCAACCCAGTCCCCTC 497

seq name: gb_est1:AA481295

seq documentation block:
LOCUS       AA481295                 480 bp    mRNA    linear   EST 14-AUG-1997
```

```
DEFINITION  zv02e05.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:746336 5',
            mRNA sequence.
ACCESSION   AA481295
VERSION     AA481295.1  GI:2210847
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 380)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 361.
FEATURES             Location/Qualifiers
     source          1. .380
                     /organism="Homo sapiens"
                     /db_xref="GDB:5945860"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:746336"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      105 a     90 c     76 g    109 t
ORIGIN


alignment_scores:
            Quality:  126.00             Length:    126
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AA481295/rev   ..

 Align seq 1/1  to reverse of: AA481295  from: 1  to: 380

     263 IleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSe 279
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     379 ATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGGTGGGACAGATTC 330

     279 rThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluT 296
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     329 CACGGAAGGATTGTTTAATATGGATGAAGATGAAGAAACTGAACATGAAA 280

     296 hrLysAlaGlyProGluArgLeuAspGlnMetValHisProValAlaGlu 312
         ||||||||||||||||||||||||||||||||||||||||||||||||||||

     279 CAAAGGCTGGTCCTGAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAG 230

     313 ArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCy 329
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     229 CGCCTGGACATCCTGATGTCTTTGGTTTTGTCCTACATGAAGGATGTCTG 180

     329 sTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgA 346
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     179 CTATGTAGATGGTAAGGTTGATAACGGCAAAACAAAGGATCTATATCGCG 130

     346 spLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCys 362
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     129 ACCTGATAAACATCTTTGACAAACTCCTGTTGCCCACCCATGCCTCCTGC 80

     363 HisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAl 379
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      79 CATGTACAGTTTTCATGTTTTACCTCTGTAGTTTCAAATTGGGATTCGC 30

     379 aGluAlaPheLeuGluHisLeuTrpLys 388
         ||||||||||||||||||||||||||
      29 AGAGGCATTTTTGCAACATCTCTGGAAA 2

seq_name : qb_est1:AA191111

seq_documentation_block:
LOCUS       AA191111                421 bp    mRNA    linear   EST 15-JAN-1997
DEFINITION  zp86a03.r1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone
            IMAGE:627052 5' similar to WP:C36E8.1 CE00909 ;, mRNA sequence.
ACCESSION   AA191111
VERSION     AA191111.1  GI:1779805
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 421)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 383.
FEATURES             Location/Qualifiers
     source          1. .421
                     /organism="Homo sapiens"
                     /db_xref="GDB:5045750"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:627052"
                     /clone_lib="Stratagene HeLa cell s3 937216"
                     /sex="female"
                     /dev_stage "HeLa S3 cell line"
                     /lab_host "SOLR (kanamycin resistant)"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: oligo dT. HeLa S3
                     epithelioid carcinoma cells grown to semi-confluency
                     without induction. Average insert size: 1.5 kb; 0.1 ZAP XR
                     Vector. -5' adaptor sequence: 5' GAATTCGGCACGAG 3'  3'
                     adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT      108 a     93 c     86 g    134 t
ORIGIN
```

```
alignment scores:
            Quality:  126.00              Length:     140
              Ratio:  0.906                 Gaps:       1
 Percent Similarity:  99.286   Percent Identity: 99.286

alignment block:
 US-09-932-678-2 x AAI91111   ..

 Align seq 1/1  to: AAI91111  from: 1  to: 421

      355 LeuLeuProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLe 371
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
        3 CTGTTGCCCACCCATGCCTCCTGCCATGTACAGTTTTTCATGTTTTACCT 52

      471 uCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpL 488
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
       53 CTGTAGTTTCAAATTGGGATTGCCAGAGGCATTTTTGGAACATCTCTGGA 102

      488 ysLysLeuGluAspProSerAsnProAlaIleIleArgGlnAlaAlaGly 404
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      103 AAAAATTGCAGGACCCAAGTAATCCTGCCATCATCAGGCAGGCTGCTGGA 152

      405 AsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIlePh 421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      153 AATTATATTGGAAGCTTTTTGGCAAGAGCTAAATTTATTCCTCTTATTAC 202

      421 rValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuA 438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      203 TGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTA 252

      438 snAsnGlnAspSer.GlyThrLysAlaPheCysAspValAlaLeuHisGl 454
          |||||||||||||| ||||||||||||||||||||||||||||||||||||
      253 ATAACCAGGATTCCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATGG 302

      454 yProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgH 471
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      303 ACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGAC 352

      471 isLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGln 487
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      353 ACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGTTTGCAGTATCTTCAC 402

      488 SerLeuAsnPheGluArg 493
          |||||||||||||||||
      403 AGTCTGAATTTTGAGCGC 420

seq_name: ab_est2:BE502966

seq_documentation_block:
LOCUS       BE502966                 490 bp    mRNA    linear   EST 04-AUG 2000
DEFINITION  hz81q07.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3214428 3'
            similar to TR:O75703 O75703 HYPOTHETICAL 12.4 KD PROTEIN. ;, mRNA
            sequence.
ACCESSION   BE502966
VERSION     BE502966.1  GI:9705374
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL. send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 427.
FEATURES             Location/Qualifiers
     source          1. .490
                     /organism="Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone="IMAGE:3214428"
                     /clone_lib "NCI_CGAP_Lu24"
                     /tissue_type "carcinoid"
                     /lab_host "DH10B"
                     /note="Organ: lung; Vector: pT7T40 Pac (Pharmacia) with a
                     modified polylinker; Plasmid DNA from the normalized
                     library NCI_CGAP_Lu5 was prepared, and ss circles were
                     made in vitro. Following HAP purification, this DNA was
                     used as tracer in a subtractive hybridization reaction.
                     The driver was PCR-amplified cDNAs from a pool of 5,000
                     clones made from the same library (cloneIDs
                     1414920-1417991 and 1520904-1522439). Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT       146 a      90 c     101 g     151 t       2 others
ORIGIN


alignment_scores:
            Quality:  125.00              Length:     125
              Ratio:  1.000                 Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x BE502966   ..

 Align seq 1/1  to: BE502966  from: 1  to: 490

       48 AsnSerProProArgLysThrValArgPheGlyGlyThrValThrGluVa 59
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
        1 AATTCTCCCCCAAGAAAAACTGTTCGGTTTGGTGGAACTGTGACAGAAGT 50

       59 IleLeuLysTyrTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysA 76
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
       51 CTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGA 100

       76 seGlnLeuLeuAspProAspIleLysAspAspGlnIleIleAsnTrpLeu 92
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      101 A CAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTG 150

       93 LeuGluPheArgSerSerIleMetTyrLeuThrLysAspPheGluGlnLe 109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      151 CTAGAATTCCGTTCTTCTATCATGTACTTGACAAAAGACTTTGAGCAACT 200

      109 uIleSerIleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrValV 126
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      201 TATCAGTATTATATTAAGATTGCCTTGGTTGAATAGAAGTCAAACAGTAG 250

      126 alGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGluThrVal 142
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      251 TGGAAGAGTATTTGGCTTTTCTTGGTAATCTTGTATCAGCACAGACTGTT 300

      143 PheLeuArgProCysLeuSerMetIleAlaSerHisPheValProProAr 159
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      301 TTCCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGTGCCTCCCCG 350

      159 gValIleIleLysGluGlyAspVal 167
          |||||||||||||||||||||||||
      351 ACTGATCATTAAGGAAGGCGATGTA 375

seq_name: ab_est2:BF203684

seq_documentation_block:
LOCUS       BF203684                 996 bp    mRNA    linear   EST 06-NOV-2000
```

```
DEFINITION  601866331F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4099398 5',
            mRNA sequence.
ACCESSION   BF203684
VERSION     BF203684.1  GI:11097270
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 996)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM967  row: n  column: 07
            High quality sequence stop: 684.
FEATURES             Location/Qualifiers
     source          1. .996
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4099398"
                     /clone_lib="NIH_MGC_17"
                     /tissue_type="rhabdomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACCAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      338 a    211 c    231 g    216 t
ORIGIN


alignment_scores:
            Quality:  122.00             Length:    122
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x BF203684   ..

 Align seg 1/1  to: BF203684  from: 1  to: 996

     430 ValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAl 446
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     155 GTTAACTGGCTGCACATATACCTTAATAACCAGGATTCGGGAACAAAGGC 204

     446 aPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaV 463
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     205 ATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTG 254

     463 alPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeu 479
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     255 TGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTG 304

     480 LysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMe 496
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     305 AAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGAT 354

     496 tSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPheP 513
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     355 GAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTT 404

     513 heAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIleIle 529
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     405 TTGCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATT 454

     530 GluArgAsnAsnArgGlnMetLeuProValIleArgSerThrAlaGlyL 546
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     455 GAGAGGAACAATCGCCAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGG 504

     546 yAspSerValGlnIle 551
         ||||||||||||||||
     505 AGACTCAGTGCAGATC 520
```

```
OM of: US-09-932-678-2 to: GenEmbl:*   out_format : pfs

Date: Jul 30, 2002  7:35 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=x11
 -Q=/cgn2_1/USPTO_spool/US09932678/runat_29072002_155155_19010/app_query.fasta_1.717
 -DB=GenEmbl -QFMT=fastap -SUFFIX=oli.rge -GAPOP=4.500
 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
 -TRANS human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
 -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
 -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09932678_#CGN1_1_4676 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-932-678-2
Query length: 651
Database: GenEmbl:*
Database sequences: 1797656
Database length: 1873333701
Search time (sec): 2262.420000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation  ..
qb_pr:AF227156          +  651.00 12690.36  0.0       2068  ! AF227156 Homo sapiens RRN3 mR
qb_pr:HSA272050         +  550.00 10711.41  0.0       3756  ! AJ272050 Homo sapiens mRNA fo
qb_pat:AX036047         +  463.00  9014.34  0.0       2040  ! AX036047 Sequence 1 from Paten
qb_pr:AK055742          +   94.00  1797.64  9.1e-92   2493  ! AK055742 Homo sapiens cDNA FL
qb_pat:AX274907         +   70.00  1331.94  7.9e-66   1461  ! AX274907 Sequence 172 from Pat
qb_pr:BC009198          +   68.00  1294.29  9.9e-64   1177  ! BC009198 Homo sapiens, Similar
qb_htg:AC040158         +   62.00  1145.32  2.0e-55   129473  ! AC040158 Homo sapiens chromo
qb_htg:AC092137         -   62.00  1144.61  2.2e-55   143900  ! AC092137 Homo sapiens chromo
qb_htg:AC007615         +   62.00  1143.26  2.6e-55   175691  ! AC007615 Homo sapiens chromo
qb_htg:AC092562         +   62.00  1142.82  2.7e-55   187643  ! AC092562 Papio hamadryas clo
qb_htg:AC009130         +   62.00  1142.68  2.8e-55   191496  ! AC009130 Homo sapiens chromo
qb_htg:AC009093         +   62.00  1142.13  3.0e-55   208008  ! AC009093 Homo sapiens chromo
qb_htg:AC097268         -   62.00  1140.96  3.4e-55   247331  ! AC097268 Pan troglodytes clo
qb_pr:BC006441          +   61.00  1152.01  8.4e-56   2627  ! BC006441 Homo sapiens, Similar
qb_pr:AC092375          -   58.00  1065.15  5.8e-51   173166  ! AC092375 Homo sapiens chromo
qb_htg:AC017077         -   58.00  1064.92  5.9e-51   179150  ! AC017077 Homo sapiens chromo
qb_pr:HUAF001549        +   58.00  1064.11  6.6e-51   202004  ! AF001549 Human Chromosome 16
qb_htg:AC092137         +   43.00   773.08  1.1e-34   143900  ! AC092137 Homo sapiens chromo
qb_pr:AC106788          +   43.00   771.94  1.2e-34   170611  ! AC106788 Homo sapiens chromo
qb_htg:AC017077         +   35.00   615.18  6.7e-26   179150  ! AC017077 Homo sapiens chromo
qb_htg:AC097269         -   35.00   614.56  7.2e-26   196413  ! AC097269 Pan troglodytes clo
qb_htg:AC009130         -   34.00   595.18  8.7e-25   191496  ! AC009130 Homo sapiens chromo
qb_htg:AC022642         +   22.00   366.55  4.7e-12   78347  ! AC022642 Homo sapiens clone R
qb_htg:AC022642         -   17.00   268.78  1.3e-06   78347  ! AC022642 Homo sapiens clone R
qb_pr:HSA343405         -   15.00   260.34  3.9e-06   824  ! AJ343405 Homo sapiens genomic s
qb_htg:AC107769         +   15.00   232.20  0.0001    53814  ! AC107769 Mus musculus clone R
qb_htg:AC107769         -   15.00   232.20  0.0001    53814  ! AC107769 Mus musculus clone R
qb_htg:AC068150         -   15.00   227.51  0.0003    108040  ! AC068150 Homo sapiens chromo
qb_htg:AC092119         -   15.00   224.38  0.0004    171930  ! AC092119 Homo sapiens chromo
qb_pr:AC092375          +   15.00   224.33  0.0004    173166  ! AC092375 Homo sapiens chromo
qb_htg:AC092285         -   15.00   220.28  0.0007    316296  ! AC092285 Homo sapiens chromo
qb_htg:AC012547         +   13.00   186.86  0.0480    135873  ! AC012547 Homo sapiens clone
qb_pr:AC013444          +   13.00   184.50  0.0649    192815  ! AC013444 Homo sapiens BAC cl
qb_pr:HS151B3F          -   11.00   189.09  0.0360    293  ! Z56854 H.sapiens CpG island DNA
qb_sts:G27379           -   11.00   186.91  0.0477    405  ! G27379 human STS SHGC-31413, se
qb_ro:AF143928          +   10.00   160.39  1.43      1140  ! AF143928 Marmota himalayana cy
qb_ro:AF143937          +   10.00   160.39  1.43      1140  ! AF143937 Marmota sibirica clon
qb_ro:AF143938          +   10.00   160.39  1.43      1140  ! AF143938 Marmota sibirica clon
qb_pl:AC003096          -   10.00   132.39  51.89     72919  ! AC003096 Arabidopsis thali
qb_pl:I518              -   10.00   130.52  65.91     96183  ! AC007060 Arabidopsis thali
qb_htg:AC095796         +   10.00   127.74  94.19     145427  ! AC095796 Rattus norvegicu
qb_pat:AX309560         -    9.00   149.98  5.44      293  ! AX309560 Sequence 2545 from
qb_ro:AF157874          +    9.00   140.83  17.57     1140  ! AF157874 Spermophilus eryth
qb_ro:AF157875          +    9.00   140.83  17.57     1140  ! AF157875 Spermophilus eryth
qb_ro:BC019693          -    9.00   139.31  21.35     1429  ! BC019693 Mus musculus, Simi

seq_name: qb_pr:AF227156

seq_documentation_block:
LOCUS       AF227156                2068 bp    mRNA    linear   PRI 29-APR-2000
DEFINITION  Homo sapiens RRN3 mRNA, complete cds.
ACCESSION   AF227156
VERSION     AF227156.1  GI:7670099
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2068)
  AUTHORS   Moorefield,B., Greene,E.A. and Reeder,R.H.
  TITLE     RNA polymerase I transcription factor Rrn3 is functionally
            conserved between yeast and human
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (9), 4724-4729 (2000)
  MEDLINE   20243763
REFERENCE   2  (bases 1 to 2068)
  AUTHORS   Moorefield,B., Greene,E.A. and Reeder,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2000) Basic Sciences, Fred Hutchinson Cancer
            Research Center, 1100 Fairview Avenue N., Seattle, WA 98109, USA
FEATURES             Location/Qualifiers
     source          1. .2068
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16p12"
     CDS             69. .2024
                     /note="similar to yeast RNA polymerase I transcription
                     factor Rrn3"
                     /codon_start=1
                     /product="RRN3"
                     /protein_id="AAF66160.1"
                     /db_xref="GI:7670100"
                     /translation="MAAPLLHTRLPGDAAASSSAVKKLGASKTGISNMRALENDFFNS
                     PPRKTVRFGGTVFEVLLKYKKGETNDFELLKNQLLDPDIKDDQIINWLLEFRSSIMYL
                     TKDFEQLISILRLPWLNRSQTVVEEYLAFLGNLVSAQTVFLRPCLSMIASHFVPPRV
                     IIKEGDVDVSDSDDEDDNLPANFDTCHRALQIIARYVPSTPWFLMPILVEKFPFVKKS
                     ERTLECYVHNLLRISVYFPTLRHEILELIIEKLLKLDVNASRQGIEDAEETATQTCGG
                     TDSTEGLFNMDEDEETEHETKAGPERLDQMVHPVAERLDILMSLVLSYMKDVCYVDGK
                     VDNGKIKDLYRDLINIFDKLLLPTHASCHVQFFMFYLCSFKLGFAEAFLEHLWKKLQD
                     PSNPAIIRQAAGNYICSFLARAKFIPLITVKSCLDLLVNWLHIYLNNQDSGTKAFCDV
                     ALHGPFYSACQAVFYTFVFRHKQLLSGNLKEGLQYLQSLNFERIVMSQLNPLKICLPS
                     VVNFFAAITNKYQLVFCYTIIERNNRQMLPVIRSTAGGDSVQICTNPLDIFFPFDPCV
                     LKRSKKFIDPIYQVWEDMSAEELQEFKKPMKKDIVEDEDDDFLKGEVPQNDTVIGIIP
                     SSFDTHFRSPSSSVGSPPVLYMQPSPL"
BASE COUNT       573 a    437 c    485 g    573 t
ORIGIN


alignment_scores:
            Quality:  651.00              Length:    651
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AF227156   ..

 Align seq 1/1  to: AF227156  from: 1  to: 2068

       1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      61 ATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGATGCGGCCGCTTC 118
```

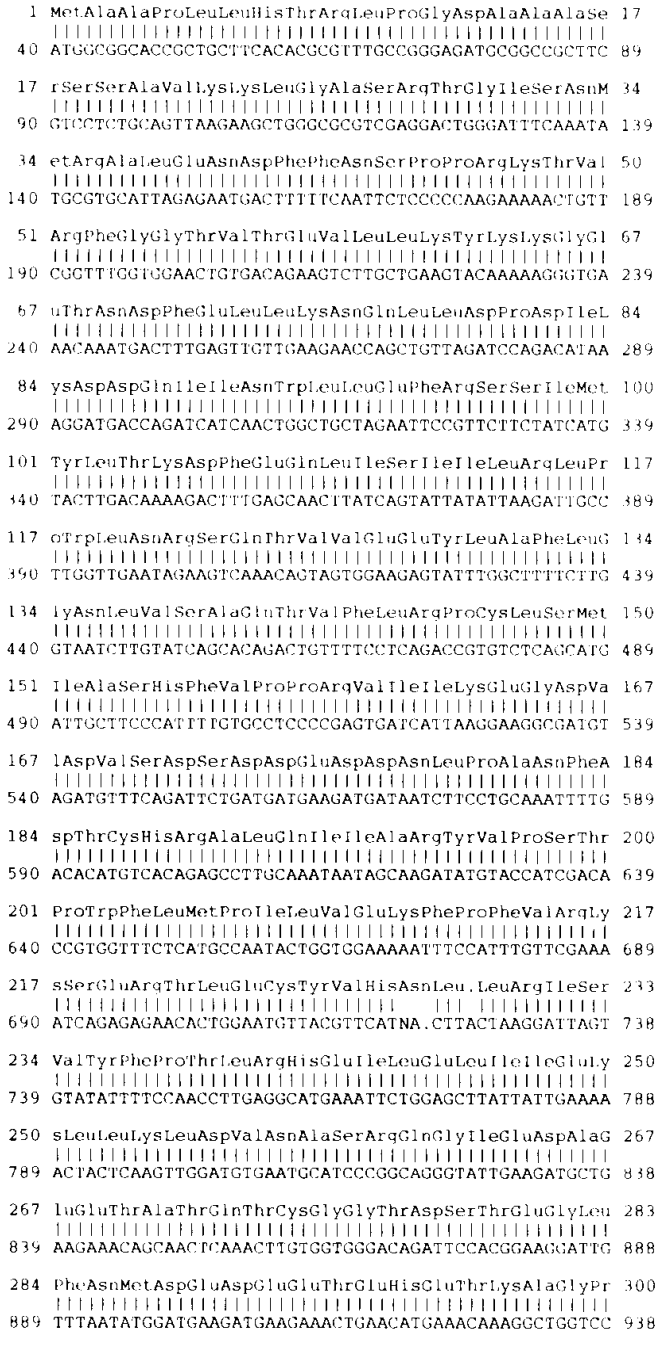

17 rSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnM 34
119 GTCCTCTGCAGTTAAGAACTGGGCGCGTCGAGGACTGGATTCAAATA 168
34 etArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrVal 50
169 TGCGTGCATTAGAGAATGACTTTTTCAATTCTCGCCCAAGAAAAACTGTT 218
51 ArgPheGlyGlyThrValPhrGluValLeuLeuLysTyrLysLysGlyGl 67
219 CGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGA 268
67 uThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleL 84
269 AACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAA 318
84 ysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100
319 AGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATG 368
101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuPr 117
369 TACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCC 418
117 eTrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLeuG 134
419 TTGGTTGAATAGAAGTCAAACAGTAGTGGAAGATTATTGGCTTTTCTTC 468
134 TyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMet 150
469 GTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATG 518
151 IleAlaSerHisPheValProProArgValIleIleLysGluGlyAspVa 167
519 ATTGCTTCCCATTTTGTGCCTCCCGAGTGATCATTAAGGAAGGCGATGT 568
167 lAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPheA 184
569 ACATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCTGCAAATTTTG 618
184 spThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr 200
619 ACACATGTCACAGACGCTTGCAAATAATAGCAAGATATGTACCATCGACA 668
201 ProTrpPheLeuMetProIleLeuValGluLysProProPheValArgLy 217
669 CCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTGTTCGAAA 718
217 sSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerV 234
719 ATCAGAGAACACTGGAATGTTACGTTCATAACTTACTAAGGATTAGTG 768
234 alTyrPheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLys 250
769 TATATTTCCAACCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAAA 818
251 LeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGl 267
819 CTACTCAAGTTGGATGTGAATGCATCCCGGCAGGGTATTGAAGATGCTGA 868
267 uGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuP 284
869 AGAAACAGCAACTCAAACTTGTGGTGCGACACGATTCCACGGAAGGATTGT 918
284 heAsnMetAspGluAspGlyGluThrGluHisGluThrLysAlaGlyPro 300
919 TTAATATGGATGAAGATGAAGAAACTGAACATGAAACAAAGGCTGGTCCT 968
301 GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleIe 317
969 GAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCGCCTGGACATCCT 1018


317 uMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyL 334
1019 GATGTCTTTGGTTTTGTCCTACATGAAGGATGTCTGCTATGTAGATGGTA 1068
334 ysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIle 350
1069 AGGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATC 1118
351 PheAspLysLeuLeuProThrHisAlaSerCysHisValGluPhePh 367
1119 TTTGACAAACTCCTGTTGCCCACCCATGCCTCCTGCCATGTACAGTTTT 1168
367 eMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuG 384
1169 CATGTTTACCTCTGTAGTTTCAAATTGGGATTCGCAGAGGCATTTTTGG 1218
384 lyHisLeuTrpLysLysLeuGluAspProSerAsnProAlaIleIleArg 400
1219 AACATCTCTGGAAAAAATTGCAGGACCCAAGTAATCCTGCCATCATCAGG 1268
401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIl 417
1269 CAGGCTGCTGGAAATTATATTGGCAAGCTTTTTGGCAAGAGCTAAATTTAT 1318
417 eProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuB 434
1319 TCCTCTAATTACTCTAAAATCATGCCTAGATCTTTGGTTAACTGGCTGC 1368
435 rsIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspVal 450
1369 ACATATACCTAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTT 1418
451 AlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPh 467
1419 GTTCTCCATGGACCATTTACTCAGCCTGCCAAGCTGTGTTCTACACCTT 1468
467 eValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuG 484
1469 CTCTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGTTTGC 1518
484 lyTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn 500
1519 ACTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAAT 1568
501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIlePh 517
1569 CCCCTGAAGATTTCCCTGCCCTCACTGGTTAACTTTTTGCTGCAATCAC 1618
517 rAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnA 534
1619 AAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAATC 1668
534 rgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGln 550
1669 GCCAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACACTCAGTGCAG 1718
551 IleCysThrAsnProLeuAspThrPhePhePheAspProCysValLe 567
1719 ATCTGCACAAACCCGCTGGACACCTTCTTCCCCTTTGATCCCTGTGTGCT 1768
567 uLysArgSerLysLysPheIleAspProIleTyrGlnValTrpGluAspM 584
1769 GAAGAGCTCAAAGAAATTCATTGATCCTATTTATCAGGTATGGGAAGACA 1818
584 etSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle 600
1819 TAGTGCTGAAGAGTTACAGGAGTTCAAGAAACCCATGAAAAAGGACATA 1868
601 ValGluAspGluAspAspAspPheLeuLysGlyGluValProGluAsnAs 617
1869 GTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCAGAATGA 1918
617 pThrValIleGlyIleThrProSerSerPheAspThrHisPheArgSerP 634

```
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
   1919 TACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAGTC 1968

    634 roSerSerSerValGlySerProProValLeuTyrMetGlnProSerPro 650
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1969 CTTCAAGTAGTGTGGGCTCCCCACCCGTGTTGTACATGCAACCCAGTCCC 2018

    651 Leu 651
        |||
   2019 CTC 2021

seq_name: qb_pr:HSA272050

seq_documentation_block:
LOCUS       HSA272050               3756 bp    mRNA    linear   PRI 07-SEP-2000
DEFINITION  Homo sapiens mRNA for transcription initiation factor IA protein
            (TIF-IA gene).
ACCESSION   AJ272050
VERSION     AJ272050.1  GI:10046713
KEYWORDS    TIF-IA; transcription initiation factor.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3756)
  AUTHORS   Bodem,J., Hoffmann-Rohrer,U., Ross,W., Delius,H., Vingron,M. and
            Grummt,I.
  TITLE     Cloning and functional characterization of transcription initiation
            factor TIF-IA, a growth-dependent regulator of ribosomal RNA
            synthesis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3756)
  AUTHORS   Bodem,J.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2000) Bodem J.J., Molekulare Biologie der Zelle
            II, Deutsches Krebsforschungszentrum, INF.280, 69120 Heidelberg,
            GERMANY
FEATURES             Location/Qualifiers
     source          1. .3756
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            23. .3756
                     /gene="TIF-IA"
     CDS             23. .1978
                     /gene="TIF-IA"
                     /function="transcription initiation factor"
                     /codon_start=1
                     /evidence=experimental
                     /product="transcription initiation factor IA protein"
                     /protein_id="CAC07955.1"
                     /db_xref="GI:10046714"
                     /translation="MAAPLLHTRLPGDAAASSSAVKKLGASRTGISNMRALENDFFNS
                     PPRKTVRFGGTVTEVLLKYKKGETNDFELLKNQLLDPDIKDDQIINWLLEFRSSIMYL
                     TKDFEQLISIILRLPWLNRSQTVVEEYLAFLGNLVSAQTVFLRPCLSMIASHFVPPRV
                     LIKEGDVDVSDSDDEDDNLPANFDTCHRALQIIARYVPSTPWFLMPILVEKFPFVRKS
                     ERTLECYVHNLLRISVYFPTLRHEILELIIEKLLKLDVNASRQGIEDAEETATQTCGG
                     TDSTEGLFNMDEDEETEHETKAGPERLDQMVHPVAERLDILMSLVLSYMKDVCYVDGK
                     VDNGKTKDLYRDLINIFDKLLLPTHASCHVQFFMFYLCSFKLGFAEAFLEHLWKKLQD
                     PSNPAIIRQAAGNYIGSFLARAKFIPLITVKSCLDLLVNWLHIYLNNQDSGTKAFCDV
                     ALHGPFYSACQAVFYXFVFRHKQLLSGNLKEGLQYLQSLNFERIVMSQINPLKICLPS
                     VVNFFAAITNKYQLVFCYTIIERNNRQMLPVIRSTAGGDSVQICTNPLDTFFPFDPCV
                     LKRSKKFIDPIYQVWEDMSAEELQEFKKPMKKDIVEDEDDDFLKGEVPQNDTVIGITP
                     SSFDTHFRSPSSSVGSPPVLYMQPSPL"
     3'UTR           1979. .3756
                     /gene="TIF-IA"
                     /evidence=experimental
BASE COUNT     1060 a     692 c     844 g    1156 t       4 others
ORIGIN


alignment_scores:
            Quality:  550.00            Length:    651
              Ratio:  0.846               Gaps:      0
```

```
 Percent Similarity:  99.846   Percent Identity: 99.846

alignment_block:
 US-09-932-678-2 x HSA272050   ..

 Align seq 1/1  to: HSA272050  from: 1  to: 3756

       1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSe 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      23 ATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGATGCGGCCGCTTC 72

      17 rSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnM 34
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      73 GTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATA 122

      34 etArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrVal 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     123 TGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTT 172

      51 ArgPheGlyGlyThrValThrGluValLeuLysTyrLysLysGlyGl 67
         |||||||||||||||||||||||||||||||||||||||||||||||||
     173 CGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGA 222

      67 uThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     223 AACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAA 272

      84 ysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     273 AGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATG 322

     101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     323 TACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCC 372

     117 oTrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLeuG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     373 TTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTG 422

     134 lyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMet 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     423 GTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCCTGTCTCAGCATG 472

     151 IleAlaSerHisPheValProProArgValIleIleLysGluGlyAspVa 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     473 ATTGCTTCCCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCGATGT 522

     167 lAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPheA 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     523 AGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCTGCAAATTTTG 572

     184 spThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     573 ACACATGTCACAGAGCCTTGCAAATAATAGCAAGATATGTACCATCGACA 622

     201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLy 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     623 CCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTCGAAA 672

     217 sSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerV 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     673 ATCAGAGAGAACACTGGAATGTTACGTTCATAACTTACTAAGGATTAGTG 722

     234 alTyrPheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLys 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     723 TATATTTTCCAACCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAAA 772

     251 LeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGl 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     773 CTACTCAAGTTGGATGTGAATGCATCCCGGCAGGGTATTGAAGATGCTGA 822

     267 uGluThrAlaThrGluThrCysGlyGlyThrAspSerThrGluGlyLeuP 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
  823 AGAAACAGCAACTCAAACTTGTGGTGGGACAGATTCCACGGAAGGATTGT 872

  284 hcAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyPro 300
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  873 TTAATATGGATGAAGATGAAGAAACTGAACATGAAACAAAGGCTGGTCCT 922

  301 GluArgLeuAspGluMetValHisProValAlaGluArgLeuAspIleLe 317
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  923 GAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCGCCTGGACATCCT 972

  417 uMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyL 434
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  973 GATGTCTTTGGTTTTGTCCTACATGAAGGATGTCTGCTATGTAGATGGTA 1022

  434 ysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIle 450
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1023 AGGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATC 1072

  451 PheAspLysLeuLeuProThrHisAlaSerCysHisValGlnPhePh 467
      ||||||||||||||||||||||||||||||||||||||||||||||||
 1073 TTTGACAAACTCCTGTTGCCCACCCATGCCTCCTGCCATGTACAGTTTT 1122

  467 eMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuG 484
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1123 CATGTTTTACCTCTGTAGTTTCAAATTGGGATTCGCAGAGGCATTTTTGG 1172

  484 EuHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArq 400
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1173 AACATCTCTGGAAAAAATTGCAGGACCCAAGTAATCCTGCCATCATCAGG 1222

  401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArqAlaLysPheIl 417
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1223 CAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAGAGCTAAATTTAT 1272

  417 eProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuH 434
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1273 TCCTCTTATTACTGTAAAATCATGCCTAGATCTTTTGCTTAACTGGCTGC 1322

  434 isIleTyrLeuAsnAsnGluAspSerGlyThrLysAlaPheCysAspVal 450
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1323 ACATATACCTTAATAACGAGGATTCGGGAACAAAGGCATTCTGCGATGTT 1372

  451 AlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPh 467
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1373 GCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTT 1422

  467 eValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuG 484
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1423 TGTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGTTTGC 1472

  484 luTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn 500
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1473 AGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAAT 1522

  501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleTh 517
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1523 CCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGCAATCAC 1572

  517 rAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArqAsnAsnA 534
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1573 AAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAATC 1622

  534 rgGlnMetLeuProValIleArqSerThrAlaGlyGlyAspSerValGln 550
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1623 GCCAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAG 1672

  551 IleCysThrAsnProLeuAspThrPhePhePro PheAspProCysValIle 567
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1673 ATCTGCACAAACCCGCTGGACACCTTCTTCCCCTTTGATCCCTGTGTGCT 1722

  567 uLysArqSerLysLysPheIleAspProIleTyrGlnValTrpGluAspM 584
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1723 GAAGAGGTCAAAGAAATTCATTGATCCTATTTATCAGGTGTGGCAAGACA 1772

  584 etSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle 600
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1773 TGAGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAAGGACATA 1822

  601 ValGluAspGluAspAspAspPheLeuLysGlyGluValProGlnAsnAs 617
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1823 GTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCAGAATGA 1872

  617 pThrValIleGlyIleThrProSerSerPheAspThrHisPheArqSerP 634
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1873 TACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTCCGAAGTC 1922

  634 roSerSerSerValGlySerProProValLeuTyrMetGlnProSerPro 650
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1923 CTTCAAGTAGTGTGGGCTCCCCCACCCGTGTTGTACATGCAACCCAGTCCC 1972

  651 Leu 651
      |||
 1973 CTC 1975

seq_name: pb_pat:AX036047

seq_documentation block:
LOCUS       AX036047               2040 bp    DNA     linear   PAT 15-NOV-2000
DEFINITION  Sequence 1 from Patent WO0055416.
ACCESSION   AX036047
VERSION     AX036047.1  GI:11191586
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2040)
  AUTHORS   Grummt,I. and Vingron,M.
  TITLE     Rna polymerase i transcription factor tif-ia
  JOURNAL   Patent: WO 0055316-A 1 21-SEP-2000;
            DEUTSCHES KREBSFORSCH (DE) ; GRUMMT INGRID (DE) ; VINGRON MARTIN
            (DE)
FEATURES             Location/Qualifiers
     source          1..2040
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             40..1995
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC16268.1"
                     /db_xref="GI:11191587"
                     /translation="MAAPLLBTRLLPGDAAASSSAVKKLGASRIGISNMRALENDFFNS
                     PPRKTVRFGGIVIEVLLKYKKGETNDFELLKNQLLDFDIKDDQTINWLLEFRSSIMYL
                     TKDFEQLISTLERLPWLNRSQTVVEEYLAFLGNLVSAQIVFLRQCLSMIASHFVPFRV
                     IIKEGDVDVSDSDDEDDNLPANFDTCHRALQLIARYVPSTPWFLMPILVEKFPFVRKS
                     ERTLECYVHKLLRISVYFPTLRHETLELIIEKLLKLDVNASRQGIEDAEETATQTCGG
                     TDSTEGLFNMDEDEETEHETKAGPERLDQMVHPVAERLDILMSLVLSYMKDVCYVDGK
                     VDNGKIKDLYRDLINIFDKLLLPTHASCHVQFFMPYLCSFKLGFAFAFLEHLWKKLQD
                     PSNPAIIRQAAGNYIGSFLARAKFIPLITVKSCLDLLVNWLHIYLNNQDSGIKAFCDV
                     ALHGPFYSACQAVFYXFVFRHKQLLSGNLKEGLQYLQSLNFERIVMSQLNPLKICLPS
                     VVNFFAAITNKYQLVFCYTIIERNNRQMLPVIRSFAGGDSVQICTNPLDTFFPFDPCV
                     LKRSKKFIDPIYQVWEDMSAEELQEFKKPMKKDIVEDEDDDFLKGEVPQNDIVIGITP
                     SSFDTHFRSPSSSVGSPPVLYMQPSPL"
BASE COUNT       563 a    433 c    474 g    568 t       2 others
ORIGIN


alignment_scores:
            Quality:  463.00              Length:    652
              Ratio:  0.713                 Gaps:      1
 Percent Similarity:  99.540   Percent Identity: 99.540

alignment_block:
 US-09-932-678-2 x AX036047   ..

  Align seq 1/1  to: AX036047  from: 1  to: 2040
```


1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSe 17
40 ATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGATGCGGCCGCTTC 89
17 rSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnM 34
90 GTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATA 139
34 etArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrVal 50
140 TGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTT 189
51 ArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGlyGl 67
190 CGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGA 239
67 uThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleL 84
240 AACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAA 289
84 ysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100
290 AGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATG 339
101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuPr 117
340 TACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCC 389
117 oTrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLeuG 134
390 TTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTG 439
134 lyAsnLeuValSerAlaGluThrValPheLeuArgProCysLeuSerMet 150
440 GTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATG 489
151 IleAlaSerHisPheValProProArgValIleIleLysGluGlyAspVa 167
490 ATTGCTTCCCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCGATGT 539
167 lAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPheA 184
540 AGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCTGCAAATTTTG 589
184 spThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr 200
590 ACACATGTCACAGAGCCTTGCAAATAATAGCAAGATATGTACCATCGACA 639
201 ProTrpPheLeuMetProIleLeuValGluLysProProPheValArgLy 217
640 CCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTCGAAA 689
217 sSerGluArgThrLeuGluCysTyrValHisAsnLeu.LeuArgIleSer 233
690 ATCAGAGAGAACACTGGAATGTTACGTTCATNA.CTTACTAAGGATTAGT 738
234 ValTyrPheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLy 250
739 GTATATTTTCCAACCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAA 788
250 sLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaG 267
789 ACTACTCAAGTTGGATGTGAATGCATCCCGGCAGGGTATTGAAGATGCTG 838
267 luGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeu 283
839 AAGAAACAGCAACTCAAACTTGTGGTGGGACAGATTCCACGGAAGGATTG 888
284 PheAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyPr 300
889 TTTAATATGGATGAAGATGAAGAAACTGAACATGAAACAAAGGCTGGTCC 938
300 oGluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleL 317
939 PGAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCGCCTGGACATCC 988
317 euMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspGly 333
989 TGATGTCTTTGGTTTTGTCCTACATGAAGGATGTCTGCTATGTAGATGGT 1038
334 LysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIl 350
1039 AAGGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACAT 1088
350 ePheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPheP 367
1089 CTTTGACAAACTCCTGTTGCCCACCCATGCCTCCTGCCATGTACAGTTTT 1138
367 heMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeu 383
1139 TCATGTTTTACCTCTGTAGTTTCAAATTGGGATTCGCAGAGGCATTTTTG 1188
384 GluHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleAr 400
1189 GAACATCTCTGGAAAAAATTGCAGGACCCAAGTAATCCTGCCATCATCAG 1238
400 gGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheI 417
1239 GCAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAGAGCTAAATTTA 1288
417 leProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeu 433
1289 TTCCTCTTATTACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTG 1338
434 HisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspVa 450
1339 CACATATACCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGT 1388
450 lAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrP 467
1389 TGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACNCCT 1438
467 heValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeu 483
1439 TTGTTTTTAGACACAAGCAGCTTTGAGCGGAAACCTGAAAGAAGGTTTG 1488
484 GlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAs 500
1489 CAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAA 1538
500 nProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleT 517
1539 TCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGCAATCA 1588
517 hrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsn 533
1589 CAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAAT 1638
534 ArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGl 550
1639 CGCCAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCA 1688
550 nIleCysThrAsnProLeuAspThrPhePheProPheAspProCysValL 567
1689 GATCTGCACAAACCCGCTGGACACCTTCTTCCCCTTTGATCCCTGTGTGC 1738
567 euLysArgSerLysLysPheIleAspProIleTyrGlnValTrpGluAsp 583
1739 TGAAGAGGTCAAAGAAATTCATTGATCCTATTTATCAGGTGTGGGAAGAC 1788
584 MetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIl 600
1789 ATGAGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAAGGACAT 1838
600 eValGluAspGluAspAspAspPheLeuLysGlyGluValProGlnAsnA 617

```
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1839 AGTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCAGAATG 1888

     617 spThrValIleGlyIleThrProSerSerPheAspThrHisPheArgSer 633
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1889 ATACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAGT 1938

     634 ProSerSerSerValGlySerProProValLeuTyrMetGlnProSerPr 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1939 CCTTCAAGTAGTGTGGGCTCCCCACCCGTGTTGTACATGCAACCCAGTCC 1988

     650 oLeu 651
         ||||
    1989 CCTC 1992

seq_name: gb_pr:AK055742

seq_documentation_block:
LOCUS       AK055742                2493 bp    mRNA    linear   PRI 31-OCT-2001
DEFINITION  Homo sapiens cDNA FLJ31180 fis, clone KIDNE2000266.
ACCESSION   AK055742
VERSION     AK055742.1  GI:16550546
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens kidney cDNA to mRNA, clone_lib:KIDNE2
            clone:KIDNE2000266.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,
            Ono,Y., Hotuta,T., Hiraoka,S., Murakawa,K., Takiguchi,S.,
            Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,
            Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,
            Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y.,
            Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y.,
            Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2493)
  AUTHORS   Isogai,T., Otsuki,T. and Sugiyama,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: RAB and
            HRI.
FEATURES             Location/Qualifiers
     source          1..2493
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="KIDNE2000266"
                     /tissue_type="kidney"
                     /clone_lib="KIDNE2"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      554 a    700 c    761 g    478 t
ORIGIN


alignment_scores:
           Quality:    94.00              Length:      94
             Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AK055742   ..

 Align seq 1/1  to: AK055742  from: 1  to: 2493

     450 ValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrTh 466
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     931 GTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACAC 980

     466 rPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyL 483
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     981 CTTCGTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGTT 1030

     483 euGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeu 499
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1031 TGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTA 1080

     500 AsnProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIl 516
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1081 AATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGCAAT 1130

     516 eThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnA 533
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1131 CACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATCGAGAGGAACA 1180

     533 snArgGlnMetLeuProValIleArgSerThr 543
         ||||||||||||||||||||||||||||||||
    1181 ATCGCCAGATGCTGCCAGTCATTAGGAGTACC 1212

seq_name: gb_pat:AX274907

seq_documentation_block:
LOCUS       AX274907                1461 bp    DNA     linear   PAT 29-OCT-2001
DEFINITION  Sequence 172 from Patent WO0172777.
ACCESSION   AX274907
VERSION     AX274907.1  GI:16547539
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Hillman,J.L., Baughn,M.R., Yue,H., Lal,P., Lu,D.A., Patterson,C.,
            Azimzai,Y., Bandman,O., Tang,Y.T., Mathur,P., Shah,P., Au-Young,J.
            and Reddy,R.
  TITLE     Transcription factors
  JOURNAL   Patent: WO 0172777-A 172 04-OCT-2001;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..1461
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 2646274CB1"
BASE COUNT      413 a    335 c    298 g    415 t
ORIGIN


alignment_scores:
           Quality:    70.00              Length:      70
             Ratio:    1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AX274907   ..

 Align seq 1/1  to: AX274907  from: 1  to: 1461

     201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLy 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     281 CCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTCGAAA 330

     217 sSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerV 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     331 ATCAGAGAGAACACTGGAATGTTACGTTCATAACTTACTAAGGATTAGTG 380
```

```
     234 alTyrPheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLys 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     381 TATATTTTCCAACCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAAA 430

     251 LeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGl 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     431 CTACTCAAGCTGGATGTGAATGCATCCCGGCAGGGTATTGAAGATGCTGA 480

     267 uGluThrAla 270
         ||||||||||
     481 AGAAACAGCA 490

seq_name: qb_pr:BC009198

seq_documentation_block:
LOCUS       BC009198                1177 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, Similar to RNA polymerase I transcription factor
            RRN3, clone MGC:15321 IMAGE:3678732, mRNA, complete cds.
ACCESSION   BC009198
VERSION     BC009198.1  GI:14327947
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1177)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 24 Row: k Column: 14
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 3355302.
FEATURES             Location/Qualifiers
     source          1. .1177
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:15321 IMAGE:3678732"
                     /tissue_type="Lymph, Burkitt lymphoma"
                     /clone_lib="NIH_MGC_8"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             316. .636
                     /codon_start=1
                     /product="Similar to RNA polymerase I transcription factor
                     RRN3"
                     /protein_id="AAH09198.1"
                     /db_xref="GI:14327948"
                     /translation="MRALENDFFNSPPRKTVRFGGTVTEVLLKYKKGETNDFELLKNQ
                     LLDPDIKDDQIINWLLEFRSSVMYLTKDFEQLISIILRLPWLNRSQTVVEEYLAFLGN
                     LVSA"
BASE COUNT      339 a    222 c    253 g    363 t
ORIGIN


alignment_scores:
            Quality:   68.00            Length:     68
            Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x BC009198   ..

 Align seq 1/1  to: BC009198  from: 1  to: 1177

      31 IleSerAsnMetArgAlaLeuGluAsnAspPhePheAsnSerProProAr 47
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     307 ATTTCAAATATGCGTGCATTAGAGAATGATTTTTTCAATTCTCCCCCAAG 456

      47 qLysThrValArgPheGlyGlyThrValThrGluValLeuLeuLysLysL 64
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     357 AAAAACTGTTCCGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACA 406

      64 ysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAsp 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     407 AAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGAT 456

      81 ProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSe 97
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     457 CCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTC 506

      97 rSer 98
         ||||
     507 TCCT 510

seq_name: qb_htg:AC040158

seq_documentation_block:
LOCUS       AC040158               129473 bp    DNA     linear   HTG 03-SEP-2000
DEFINITION  Homo sapiens chromosome 16 clone CTA-133B4, WORKING DRAFT SEQUENCE,
            14 unordered pieces.
ACCESSION   AC040158
VERSION     AC040158.3  GI:9965546
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 129473)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 129473)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Sep 3, 2000 this sequence version replaced gi:9090652.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 0
            Center clone name: CIT978SKA_133B4
            -------
            Summary Statistics
            Consensus quality: 90046 bases at least Q40
            Consensus quality: 99266 bases at least Q30
            Consensus quality: 104098 bases at least Q20
            Estimated insert size: 125000; agarose-fp estimation
            Estimated insert size: 128173; sum-of-contigs estimation
            Quality coverage:  4.51 in Q20 bases; agarose-fp estimation
            Quality coverage:  4.4 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
```

```
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      1718: contig of 1718 bp in length
            *     1719      1818: gap of unknown length
            *     1819      3832: contig of 2014 bp in length
            *     3833      3932: gap of unknown length
            *     3933      5567: contig of 1635 bp in length
            *     5568      5667: gap of unknown length
            *     5668      7520: contig of 1853 bp in length
            *     7521      7620: gap of unknown length
            *     7621      9146: contig of 1526 bp in length
            *     9147      9246: gap of unknown length
            *     9247     11144: contig of 1898 bp in length
            *    11145     11244: gap of unknown length
            *    11245     12746: contig of 1502 bp in length
            *    12747     12846: gap of unknown length
            *    12847     14568: contig of 1722 bp in length
            *    14569     14668: gap of unknown length
            *    14669     19396: contig of 4728 bp in length
            *    19397     19496: gap of unknown length
            *    19497     25634: contig of 6138 bp in length
            *    25635     25734: gap of unknown length
            *    25735     36780: contig of 11046 bp in length
            *    36781     36880: gap of unknown length
            *    36881     50223: contig of 13343 bp in length
            *    50224     50423: gap of unknown length
            *    50424     71412: contig of 20989 bp in length
            *    71413     71812: gap of unknown length
            *    71813    129473: contig of 57661 bp in length.
FEATURES             Location/Qualifiers
     source          1..129473
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="CTA_13384"
                     /clone_lib="CalTech human BAC library A"
BASE COUNT    30715 a   30506 c   41241 g   25543 t    1468 others
ORIGIN


alignment scores:
            Quality:    62.00              Length:      62
              Ratio:     1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-932-678-2 x AC040158   ..

 Align seq 1/1   to: AC040158   from: 1   to: 129473

      423 PheValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLe 437
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   115644 ACTGTAAAATCATGCCTAGATCTTTGCTTAACTGGCTGCACATATACCT 115692

      437 uAsnAsnGluAspSerGlyPheLysAlaPheCysAspValAlaLeuHisG 454
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   115693 TAATAACTAGGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATG 115742

      454 lyProPheTyrSerAlaCysGluAlaValPheTyrThrPheValPheArg 470
          ||||||||||||||||||||||||||||||||||||||||||||||||
   115743 GACCATTTTACTCACGCTGCCAAGCTGTGTTCTACACTTTGTTTTTAGA 115792

      471 HisLysGluLeuLeuSerGlyAsnLeuLysGluGly 482
          ||||||||||||||||||||||||||||||||||
   115793 CACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGCT 115828

Seq name: qb btq:AC092137

Seq documentation block:
LOCUS       AC092137               143900 bp    DNA     linear   HTG 22-JUN-2001
DEFINITION  Homo sapiens chromosome 16 clone RP11-344B15, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.

ACCESSION   AC092137
VERSION     AC092137.1  GI:14522983
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 143900)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 143900)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     ------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            --  -
            Project Information
            Center Project Name: 548644
            Center clone name: RPCI-11_344B15
            -  ----
            Summary Statistics
            Consensus quality: 124071 bases at least Q40
            Consensus quality: 136140 bases at least Q30
            Consensus quality: 137741 bases at least Q20
            Estimated insert size: 211000; agarose fp estimation
            Estimated insert size: 142200; sum of contigs estimation
            Quality coverage:  4.0 in Q20 bases; agarose fp estimation
            Quality coverage:  5.94 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      1033: contig of 1033 bp in length
            *     1034      1133: gap of unknown length
            *     1134      2547: contig of 1414 bp in length
            *     2548      2647: gap of unknown length
            *     2648      4205: contig of 1558 bp in length
            *     4206      4305: gap of unknown length
            *     4306      6564: contig of 2259 bp in length
            *     6565      6664: gap of unknown length
            *     6665      9065: contig of 2401 bp in length
            *     9066      9165: gap of unknown length
            *     9166     12059: contig of 2894 bp in length
            *    12060     12159: gap of unknown length
            *    12160     19295: contig of 7136 bp in length
            *    19296     19395: gap of unknown length
            *    19396     25668: contig of 6273 bp in length
            *    25669     25768: gap of unknown length
            *    25769     31013: contig of 5245 bp in length
            *    31014     31113: gap of unknown length
            *    31114     36555: contig of 5442 bp in length
            *    36556     36655: gap of unknown length
            *    36656     45338: contig of 8683 bp in length
            *    45339     45438: gap of unknown length
            *    45439     52918: contig of 7480 bp in length
            *    52919     53018: gap of unknown length
            *    53019     65720: contig of 12702 bp in length
            *    65721     65820: gap of unknown length
            *    65821     76109: contig of 10289 bp in length
            *    76110     76209: gap of unknown length
            *    76210     85439: contig of 9230 bp in length
            *    85440     85539: gap of unknown length
            *    85540    101620: contig of 16081 bp in length
            *   101621    101720: gap of unknown length
```

```
            *    101721    122163: contig of 20443 bp in length
            *    122164    122263: gap of unknown length
            *    122264    143900: contig of 21637 bp in length.
FEATURES             Location/Qualifiers
     source          1. .143900
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-344H15"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT    37791 a   34128 c   34320 g   35886 t    1775 others
ORIGIN


alignment_scores:
            Quality:    62.00             Length:      62
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AC092137/rev    ..

 Align seq 1/1  to reverse of: AC092137  from: 1  to: 143900

      421 ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLe 437
          |||||||||||||||||||||||||||||||||||||||||||||||||||
    66649 ACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCT 66600

      437 uAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisG 454
          ||||||||||||||||||||||||||||||||||||||||||||||||||
    66599 TAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATG 66550

      454 lyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArg 470
          ||||||||||||||||||||||||||||||||||||||||||||||||||
    66549 GACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGA 66500

      471 HisLysGlnLeuLeuSerGlyAsnLeuLysGluGly 482
          |||||||||||||||||||||||||||||||||||||
    66499 CACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGT 66464

seq_name: gb_htg:AC007615

seq_documentation_block:
LOCUS        AC007615              175691 bp    DNA     linear   HTG 03-JUL-2001
DEFINITION   Homo sapiens chromosome 16 clone RP11-528K16, WORKING DRAFT
             SEQUENCE, 1 ordered pieces.
ACCESSION    AC007615
VERSION      AC007615.6  GI:14589428
KEYWORDS     HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 175691)
  AUTHORS    DOE Joint Genome Institute.
  TITLE      Sequencing of Human Chromosome 16
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 175691)
  AUTHORS    Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
             Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
             Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
             Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
             McMurry,K., Han,C. and Deaven,L.
  TITLE      Direct Submission
  JOURNAL    Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint
             Genome Institute, Los Alamos National Laboratory, MS M888, Los
             Alamos, NM 87545, USA
COMMENT      On Jul 3, 2001 this sequence version replaced gi:13928651.
             Sequence Quality Assessment:
              This entry has been annotated with sequence quality
              estimates computed by the Phrap assembly program.
              All manually edited bases have been reduced to quality zero.
              Quality levels above 40 are expected to have less than
              1 error in 10,000 bp.
              Base-by-base quality values are not generally visible from the
              GenBank flat file format but are available as part
              of this entry's ASN.1 file.
             ----------.
             Sequence Quality Assessment:
              This entry has been annotated with sequence quality
              estimates computed by the Phrap assembly program.
              All manually edited bases have been reduced to quality zero.
              Quality levels above 40 are expected to have less than
              1 error in 10,000 bp.
              Base-by-base quality values are not generally visible from the
              GenBank flat file format but are available as part
              of this entry's ASN.1 file.
             ----------.
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 1 contigs. Gaps between the contigs
             * are represented as runs of N. The order of the pieces
             * is believed to be correct as given, however the sizes
             * of the gaps between them are based on estimates that have
             * provided by the submittor.
             * This sequence will be replaced
             * by the finished sequence as soon as it is available and
             * the accession number will be preserved.
             *         1    175691: contig of 175691 bp in length
FEATURES             Location/Qualifiers
     source          1. .175691
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone "RP11-528K16"
BASE COUNT    42392 a   42125 c   45277 g   45896 t       1 others
ORIGIN


alignment_scores:
            Quality:    62.00             Length:      62
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AC007615    ..

 Align seq 1/1  to: AC007615  from: 1  to: 175691

      421 ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLe 437
          |||||||||||||||||||||||||||||||||||||||||||||||||||
   31007 ACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCT 31056

      437 uAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisG 454
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   31057 TAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATG 31106

      454 lyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArg 470
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   31107 GACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGA 31156

      471 HisLysGlnLeuLeuSerGlyAsnLeuLysGluGly 482
          |||||||||||||||||||||||||||||||||||||
   31157 CACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGT 31192

seq_name: gb_htg:AC092562

seq_documentation_block:
LOCUS        AC092562              187643 bp    DNA     linear   HTG 18-OCT 2001
DEFINITION   Papio hamadryas clone RP41-285I13, WORKING DRAFT SEQUENCE, 8
             unordered pieces.
ACCESSION    AC092562
VERSION      AC092562.3  GI:16258968
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE       baboon.
  ORGANISM   Papio hamadryas
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 187643)
  AUTHORS   Muzny,D.M., Adams,C., Adio Ochola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Briova,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Dept,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Duque Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraquto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,L., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Jondah,S., Karlsson,E., Kelly,S., Khan,D., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Louiseqed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., Mcleod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newton,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Ogoh,M., Okwuonu,G., Oraqunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Soderqren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tant,H., Tansey,J., Taylor,C., Taylor,T., Teltrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 187643)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     on Oct 18, 2001 this sequence version replaced gi:15808491.
                        Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc help@bcm.tmc.edu
                    -- Project Information
            Center project name: ZUAG
            Center clone name: RP41 285I13
                    -- Summary Statistics
            Sequencing vector: Plasmid; M77789
            Chemistry: dye terminator Big Dye: 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 191459 bases at least Q40
            Consensus quality: 193431 bases at least Q30
            Consensus quality: 194801 bases at least Q20
            Estimated insert size: 188848, sum of contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-tp estimation
            Quality coverage: 8.4x in Q20 bases; sum-of contigs estimation

            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     67581: contig of 67581 bp in length
            *   67582     67681: gap of unknown length
            *   67682    113789: contig of 46108 bp in length
            *  113790    113889: gap of unknown length
            *  113890    131405: contig of 17516 bp in length
            *  131406    131505: gap of unknown length
            *  131506    144468: contig of 12963 bp in length
            *  144469    144568: gap of unknown length
            *  144569    163180: contig of 18612 bp in length
            *  163181    163280: gap of unknown length
            *  163281    176573: contig of 13293 bp in length
            *  176574    176673: gap of unknown length
            *  176674    185462: contig of 8789 bp in length
            *  185463    185562: gap of unknown length
            *  185563    187643: contig of 2081 bp in length.
FEATURES             Location/Qualifiers
     source          1..187643
                     /organism="Papio hamadryas"
                     /db_xref="taxon:9557"
                     /clone="RP41-285I13"
BASE COUNT    49173 a   41946 c   44647 g   51158 t     719 others
ORIGIN


alignment_scores:
            Quality:   62.00              Length:      62
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AC092562    ..

 Align seq 1/1  to: AC092562  from: 1  to: 187643

      421 ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrIe 437
          ||||||||||||||||||||||||||||||||||||||||||||||||||
    32590 A'TGTAAAATCGTGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCT 32639

      437 uAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisG 454
          ||||||||||||||||||||||||||||||||||||||||||||||||||
    32640 TAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATG 32689

      454 lyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArq 470
          ||||||||||||||||||||||||||||||||||||||||||||||||||
    32690 GGCCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTAGA 32739

      471 HisLysGluLeuLeuSerGlyAsnLeuLysGluGly 482
          ||||||||||||||||||||||||||||||||||||
    32740 CACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGT 32775

seq_name: gb_htg:AC009130

seq_documentation_block:
LOCUS       AC009130              191496 bp    DNA     linear   HTG 21-JUN-2000
DEFINITION  Homo sapiens chromosome 16 clone RP11-501P17, WORKING DRAFT
            SEQUENCE, 43 unordered pieces.
ACCESSION   AC009130
VERSION     AC009130.6  GI:8575965
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191496)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
```

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191496)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jun 21, 2000 this sequence version replaced gi:7689948.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            --------
            Project Information
            Center Project Name: 599126
            Center clone name: RPCI-11_501P17
            --------
            Summary Statistics
            Consensus quality: 158339 bases at least Q40
            Consensus quality: 176831 bases at least Q30
            Consensus quality: 180633 bases at least Q20
            Estimated insert size: 201840; agarose-fp estimation
            Estimated insert size: 187296; sum-of-contigs estimation
            Quality coverage:  3.4 in Q20 bases; agarose-fp estimation
            Quality coverage:  3.67 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 43 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1071: contig of 1071 bp in length
            *     1072     1171: gap of unknown length
            *     1172     2358: contig of 1187 bp in length
            *     2359     2458: gap of unknown length
            *     2459     3680: contig of 1222 bp in length
            *     3681     3780: gap of unknown length
            *     3781     5278: contig of 1498 bp in length
            *     5279     5378: gap of unknown length
            *     5379     6522: contig of 1144 bp in length
            *     6523     6622: gap of unknown length
            *     6623     8589: contig of 1967 bp in length
            *     8590     8689: gap of unknown length
            *     8690    10266: contig of 1577 bp in length
            *    10267    10366: gap of unknown length
            *    10367    11714: contig of 1348 bp in length
            *    11715    11814: gap of unknown length
            *    11815    13443: contig of 1629 bp in length
            *    13444    13543: gap of unknown length
            *    13544    14645: contig of 1102 bp in length
            *    14646    14745: gap of unknown length
            *    14746    15751: contig of 1006 bp in length
            *    15752    15851: gap of unknown length
            *    15852    16868: contig of 1017 bp in length
            *    16869    16968: gap of unknown length
            *    16969    19037: contig of 2069 bp in length
            *    19038    19137: gap of unknown length
            *    19138    20794: contig of 1657 bp in length
            *    20795    20894: gap of unknown length
            *    20895    22876: contig of 1982 bp in length
            *    22877    22976: gap of unknown length
            *    22977    24751: contig of 1775 bp in length
            *    24752    24851: gap of unknown length
            *    24852    27381: contig of 2530 bp in length
            *    27382    27481: gap of unknown length
            *    27482    30339: contig of 2858 bp in length
            *    30340    30439: gap of unknown length
            *    30440    32397: contig of 1958 bp in length
            *    32398    32497: gap of unknown length
            *    32498    34880: contig of 2383 bp in length
            *    34881    34980: gap of unknown length
            *    34981    37578: contig of 2598 bp in length
            *    37579    37678: gap of unknown length
            *    37679    40004: contig of 2326 bp in length
            *    40005    40104: gap of unknown length
            *    40105    43544: contig of 3440 bp in length
            *    43545    43644: gap of unknown length
            *    43645    46429: contig of 2785 bp in length
            *    46430    46529: gap of unknown length
            *    46530    49686: contig of 3157 bp in length
            *    49687    49786: gap of unknown length
            *    49787    52269: contig of 2483 bp in length
            *    52270    52369: gap of unknown length
            *    52370    55195: contig of 2826 bp in length
            *    55196    55295: gap of unknown length
            *    55296    58516: contig of 3221 bp in length
            *    58517    58616: gap of unknown length
            *    58617    61359: contig of 2743 bp in length
            *    61360    61459: gap of unknown length
            *    61460    66151: contig of 4692 bp in length
            *    66152    66251: gap of unknown length
            *    66252    73145: contig of 6894 bp in length
            *    73146    73245: gap of unknown length
            *    73246    78200: contig of 4955 bp in length
            *    78201    78300: gap of unknown length
            *    78301    83186: contig of 4886 bp in length
            *    83187    83286: gap of unknown length
            *    83287    88570: contig of 5284 bp in length
            *    88571    88670: gap of unknown length
            *    88671    95276: contig of 6606 bp in length
            *    95277    95376: gap of unknown length
            *    95377   102234: contig of 6858 bp in length
            *   102235   102334: gap of unknown length
            *   102335   107630: contig of 5296 bp in length
            *   107631   107730: gap of unknown length
            *   107731   120220: contig of 12490 bp in length
            *   120221   120320: gap of unknown length
            *   120321   134451: contig of 14131 bp in length
            *   134452   134551: gap of unknown length
            *   134552   144891: contig of 10340 bp in length
            *   144892   144991: gap of unknown length
            *   144992   159434: contig of 14443 bp in length
            *   159435   159534: gap of unknown length
            *   159535   175863: contig of 16329 bp in length
            *   175864   175963: gap of unknown length
            *   175964   191496: contig of 15533 bp in length.
FEATURES             Location/Qualifiers
     source          1. .191496
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RPCI-501P17"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT    47031 a  47510 c  47005 g  45732 t   4218 others
ORIGIN


alignment_scores:
            Quality:   62.00              Length:      62
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AC009130   ..

 Align seq 1/1  to: AC009130  from: 1  to: 191496

     421 ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrIle 467
         |||||||||||||||||||||||||||||||||||||||||||||||||||
   26624 ACTGTAAAATCATGCCTAGATCTTTGGTTAACTGGCTGCACATATACCT 26673

     467 uAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisG 454
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   26674 TAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATG 26723
```

```
     454 lyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArg 470
         |||||||||||||||||||||||||||||||||||||||||||||||||||
   26724 GACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGA 26773

     471 HisLysGlnLeuLeuSerGlyAsnLeuLysGluGly 482
         ||||||||||||||||||||||||||||||||||||
   26774 CACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGT 26809

seq_name: gb_htg:AC009093

seq_documentation_block:
LOCUS       AC009093                208008 bp    DNA     linear   HTG 04-SEP 2001
DEFINITION  Homo sapiens chromosome 16 clone RP11-426C22, WORKING DRAFT
            SEQUENCE, 1 ordered pieces.
ACCESSION   AC009093
VERSION     AC009093.8  GI:15426058
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 208008)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 208008)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03 AUG 1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Sep 4, 2001 this sequence version replaced gi:13786300.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base by base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base by base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
             ---  ------ --
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1  208008: contig of 208008 bp in length.
FEATURES             Location/Qualifiers
     source          1..208008
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-426C22"
BASE COUNT    49489 a  50452 c  53814 g  54252 t       1 others
ORIGIN

alignment_scores:
            Quality:    62.00             Length:      62
              Ratio:    1.000               Gaps:       0
 Percent Similarity:  100.000   Percent Identity:  100.000

alignment_block:
 US-09-932-678-2 x AC009093  ..

 Align seq 1/1  to: AC009093  from: 1  to: 208008

     421 ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLe 437
         |||||||||||||||||||||||||||||||||||||||||||||||||||
   28489 ACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCT 28538

     437 uAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisG 454
         |||||||||||||||||||||||||||||||||||||||||||||||||||
   28539 TAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATG 28588

     454 lyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArg 470
         |||||||||||||||||||||||||||||||||||||||||||||||||||
   28589 GACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGA 28638

     471 HisLysGlnLeuLeuSerGlyAsnLeuLysGluGly 482
         ||||||||||||||||||||||||||||||||||||
   28639 CACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGT 28674

seq_name: gb_htg:AC097268

seq_documentation_block:
LOCUS       AC097268                247331 bp    DNA     linear   HTG 20-OCT 2001
DEFINITION  Pan troglodytes clone RP43-4308, WORKING DRAFT SEQUENCE, 11
            unordered pieces.
ACCESSION   AC097268
VERSION     AC097268.1  GI:16117531
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      chimpanzee.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 247331)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bubay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,L.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
```

```
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 247331)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
              Center project name: ZUAV
              Center clone name: RP43-430B
            -------------- Summary Statistics
              Sequencing vector: Plasmid; M77789
              Chemistry: Dye-terminator Big Dye: 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 246206 bases at least Q40
              Consensus quality: 249607 bases at least Q30
              Consensus quality: 252129 bases at least Q20
              Estimated insert size: 250832; sum-of-contigs estimation
              Quality coverage: 0x in Q20 bases; agarose-fp estimation
              Quality coverage: 8.9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    58134: contig of 58134 bp in length
            *    58135    58234: gap of unknown length
            *    58235   104153: contig of 45919 bp in length
            *   104154   104253: gap of unknown length
            *   104254   142622: contig of 38369 bp in length
            *   142623   142722: gap of unknown length
            *   142723   169545: contig of 26823 bp in length
            *   169546   169645: gap of unknown length
            *   169646   190172: contig of 20527 bp in length
            *   190173   190272: gap of unknown length
            *   190273   203875: contig of 13603 bp in length
            *   203876   203975: gap of unknown length
            *   203976   217146: contig of 13171 bp in length
            *   217147   217246: gap of unknown length
            *   217247   232417: contig of 15171 bp in length
            *   232418   232517: gap of unknown length
            *   232518   241219: contig of 8702 bp in length
            *   241220   241419: gap of unknown length
            *   241320   245193: contig of 3874 bp in length
            *   245194   245293: gap of unknown length
            *   245294   247331: contig of 2038 bp in length.
FEATURES             Location/Qualifiers
     source          1. .247331
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="RP43-430B"
BASE COUNT    67808 a  55778 c  54999 g  67717 t   1029 others
ORIGIN


alignment_scores:
```

```
            Quality:   62.00              Length:      62
              Ratio:   1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AC097268/rev    ..

 Align seq 1/1  to reverse of: AC097268  from: 1  to: 247331

      421 ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLe 467
          ||||||||||||||||||||||||||||||||||||||||||||||||||
    72903 ACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCT 72854

      437 uAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisC 454
          ||||||||||||||||||||||||||||||||||||||||||||||||||
    72853 TAATAACCAGGATCGGCAACAAAGGCATTCTGCGATGTTGCTCTCCATG 72804

      454 lyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArg 470
          ||||||||||||||||||||||||||||||||||||||||||||||||||
    72803 GACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGA 72754

      471 HisLysGlnLeuLeuSerGlyAsnLeuLysGluGly 482
          ||||||||||||||||||||||||||||||||||||
    72753 CACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGT 72718

seq_name: qb_pr:BC006441

seq_documentation_block:
LOCUS       BC006441                2627 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, Similar to RNA polymerase I transcription factor
            RRN3, clone MGC:13169 IMAGE:3351791, mRNA, complete cds.
ACCESSION   BC006441
VERSION     BC006441.1  GI:13623642
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2627)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Gain,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 18 Row: d Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency OKF
            analysis.
FEATURES             Location/Qualifiers
     source          1. .2627
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /clone "MGC:13169 IMAGE:3451791"
                     /tissue_type="Eye, retinoblastoma"
                     /clone_lib "NIH_MGC_16"
                     /lab_host "DH10B_R"
                     /note "Vector: pCTB7"
     CDS             1698..2156
                     /codon_start=1
                     /product "Similar to RNA polymerase I transcription factor
                     RRN3"
                     /protein_id="AAH06441.1"
                     /db_xref="GI:13623643"
                     /translation="MGFAEAFLEHLWKNLQDPSNPAIIRQAAGNYIGSFLARAKFISL
                     IFVKPCLDLLVNWLRIYLNNQDSGTKAFCDVALHGPFYSACQAVFYFFVFRHKQLLSG
                     NLKECLQYFQSLNFERIVMSQLNPLKICLPSVVNFFAAITKMKTCGYGWW"
BASE COUNT       847 a    492 c    525 g    773 t
ORIGIN


alignment scores:
            Quality:   61.00              Length:      61
              Ratio:  1.000                Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-932-678-2 x BC006441   ..

 Align seq 1/1  to: BC006441  from: 1  to: 2627

      425 CysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAs 441
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1845 TGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCAGGA 1894

      441 pSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrS 458
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1895 TTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACT 1944

      458 erAlaCysGluAlaValPheTyrThrPheValPheArgHisLysGlnLeu 474
          ||||||||||||||||||||||||||||||||||||||||||||||||
     1945 CAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTT 1994

      475 LeuSerGlyAsnLeuLysGluGlyLeuGlnTyr 485
          |||||||||||||||||||||||||||||||
     1995 TTGAGCGGAAACCTGAAAGAAGGTTTGCAGTAT 2027

Seq name: gb|pr:AC092375

Seq documentation block:
LOCUS       AC092375               173166 bp    DNA     linear   PRI 02-NOV-2001
DEFINITION  Homo sapiens chromosome 16 clone RP11-645C24, complete sequence.
ACCESSION   AC092375
VERSION     AC092375.3  GI:16596541
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173166)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173166)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 173166)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Nov 2, 2001 this sequence version replaced gi:15667998.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base by base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
FEATURES             Location/Qualifiers
     source          1..173166
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-645C24"
     misc_feature    61420..63875
                     /note="unresolved tandem repeat"
BASE COUNT    51459 a   47672 c   46529 g   47506 t
ORIGIN


alignment scores:
            Quality:   58.00              Length:      58
              Ratio:  1.000                Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US-09-932-678-2 x AC092375/rev   ..

 Align seq 1/1  to reverse of: AC092375  from: 1  to: 173166

      427 CysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAs 441
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   26104 TGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCAGGA 26055

      441 pSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrS 458
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   26054 TTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACT 26005

      459 erAlaCysGluAlaValPheTyrThrPheValPheArgHisLysGlnLeu 474
          ||||||||||||||||||||||||||||||||||||||||||||||||
   26004 CAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTT 25955

      475 LeuSerGlyAsnLeuLysGluGly 482
          |||||||||||||||||||||||||
   25954 TTGAGCGGAAACCTGAAAGAAGGT 25931
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 02:40:21 ; Search time 1685.89 Seconds
(without alignments)
16556.064 Million cell updates/sec

Title: US-09-932-678-1
Perfect score: 2068
Sequence: 1 acagaggctgtggctggaag.....tgacatttgggattcccgat 2068

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	547	26.5	550	9	AL597171	AL597171 DKFZp313K
2	510	24.7	788	10	BG502871	BG502871 602550430
3	500	24.2	535	9	AA213789	AA213789 zr91a11.s
4	497	24.0	1037	10	BG428305	BG428305 602498983
5	467	22.6	563	9	AW408066	AW408066 UI-HF-BM0
6	460	22.2	733	10	BI561523	BI561523 603256222
7	458	22.1	458	10	BE549643	BE549643 7b40g08.x
8	449	21.7	671	9	AV703279	AV703279 AV703279
9	444	21.5	564	9	AW958173	AW958173 EST370243
10	415	20.1	697	10	BG611364	BG611364 602612725
11	389	18.8	440	9	AA811628	AA811628 ob74d03.s
12	385	18.6	996	10	BF203684	BF203684 601866331
c 13	380	18.4	380	9	AA481295	AA481295 zv02e05.r
14	375	18.1	490	10	BE502966	BE502966 hz81q07.x
15	373	18.0	990	10	BG167061	BG167061 602344740
16	371	17.9	422	9	BE093077	BE093077 RC5-BT074
17	357	17.3	479	9	AW239267	AW239267 xb38b04.y
18	337	16.3	498	10	BG149493	BG149493 nad29e04.
19	312	15.1	348	9	AW015483	AW015483 UI-H-BI0
20	300	14.5	421	9	AA191111	AA191111 zp86a03.r
c 21	283	13.7	507	9	AW768543	AW768543 hk62102.x
22	278	13.4	316	10	BG149669	BG149669 nad32a11.
23	273	13.2	382	9	AV649508	AV649508 AV649508
24	273	13.2	828	10	BI761160	BI761160 603043657
25	268	13.0	503	9	AI799954	AI799954 we41d02.x
26	267	12.9	632	10	BE614096	BE614096 601503834
27	264	12.8	497	10	BF221544	BF221544 7o69d12.x
28	264	12.8	498	10	BF110176	BF110176 7b50d06.x
29	261	12.6	736	10	BI464721	BI464721 603202341
30	260	12.6	793	10	BG501864	BG501864 602548963
31	257	12.4	835	10	BI601238	BI601238 603245755
32	246	11.9	461	9	AI928274	AI928274 wo67c08.x
33	232	11.2	370	9	BE166014	BE166014 MR3-HT048
34	218	10.5	258	10	N42382	N42382 yy15t11.r1
35	217	10.5	742	10	BE870692	BE870692 601448628
36	214	10.3	529	9	AW968124	AW968124 EST480319
c 37	209	10.1	448	9	AA213873	AA213873 zr91a11.r
38	209	10.1	727	10	BI549394	BI549394 603190049
39	205	9.9	946	10	BG535886	BG535886 602564040
40	203	9.8	356	10	BE832086	BE832086 PM0-MT007
41	197	9.5	253	9	AA376316	AA376316 EST88717
42	197	9.5	278	10	BF215433	BF215433 601880964
c 43	196	9.5	406	10	BI018122	BI018122 CM1-MT024
44	196	9.5	505	10	BG532530	BG532530 602562084
c 45	193	9.3	327	9	AW802818	AW802818 IL2-UM007

ALIGNMENTS

RESULT 1
AL597171
LOCUS AL597171 550 bp mRNA linear EST 14 AUG-2001
DEFINITION DKFZp313K0411_r1 313 (synonym: hlcc2) Homo sapiens cDNA clone DKFZp313K0411 5', mRNA sequence.
ACCESSION AL597171
VERSION AL597171.1 GI:15154983
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 550)
AUTHORS Koehrer,K., Beyer,A., Mewes,W., Weil,B. and Wiemann,S.
TITLE EST (Koehrer,K., Beyer,A., Mewes,H.W., Weil,B. and Wiemann,S.)
JOURNAL Unpublished (1999)
COMMENT Contact: Koehrer K
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Charite, Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
No s1 sequence available.
This clone (DKFZp313K0411) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES Location/Qualifiers
source 1..550
/organism "Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp313K0411"
/clone_lib="313 (synonym: hlcc2)"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB; cDNA-collection"
BASE COUNT 140 a 116 c 140 g 153 t 1 others

```
ORIGIN

  Query Match             26.5%;  Score 547;  DB 9;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-272;
  Matches  547;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         23 gctgggcatccggcctgaggcgcagcggtcgcgttagttcggcccaatggcggcaccgct 82
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GCTGGGCATCCGGCCTGAGGCGCAGCGGTCGCGTTAGTTCGGCCCAATGGCGGCACCGCT 61

Qy         83 gcttcacacgcgtttgccgggagatgcggccgcttcgtcctctgcagttaagaagctggg 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GCTTCACACGCGTTTGCCGGGAGATGCGGCCGCTTCGTCCTCTGCAGTTAAGAAGCTGGG 121

Qy        143 cgcgtcgaggactgggatttcaaatatgcgtgcattagagaatgactttttcaattctcc 202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 CGCGTCGAGGACTGGGATTTCAAATATGCGTGCATTAGAGAATGACTTTTTCAATTCTCC 181

Qy        203 cccaagaaaaactgttcggtttggtggaactgtgacagaagtcttgctgaagtacaaaaa 262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 CCCAAGAAAAACTGTTCGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAA 241

Qy        263 gggtgaaacaaatgactttgagttgttgaagaaccagctgttagatccagacataaagga 322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 GGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAAAGGA 301

Qy        323 tgaccagatcatcaactggctgctagaattccgttcttctatcatgtacttgacaaaaga 382
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 TGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATGTACTTGACAAAAGA 361

Qy        383 ctttgagcaacttatcagtattatattaagattgccttggttgaatagaagtcaaacagt 442
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 CTTTGAGCAACTTATCAGTATTATATTAAGATTGCCTTGGTTGAATAGAAGTCAAACAGT 421

Qy        443 agtggaagagtatttggcttttcttggtaatcttgtatcagcacagactgttttcctcag 502
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 AGTGGAAGAGTATTTGGCTTTTCTTGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAG 481

Qy        503 accgtgtctcagcatgattgcttcccatttgtgcctccccgagtgatcattaaggaagg 562
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 ACCGTGTCTCAGCATGATTGCTTCCCATTTGTGCCTCCCCGAGTGATCATTAAGGAAGG 541

Qy        563 cgatgta 569
              |||||||
Db        542 CGATGTA 548


RESULT      2
BG502871
LOCUS       BG502871                 788 bp    mRNA    linear   EST 27-MAR-2001
DEFINITION  602550430F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4658071 5',
            mRNA sequence.
ACCESSION   BG502871
VERSION     BG502871.1  GI:13464388
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 788)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1451  row: 1  column: 08
            High quality sequence stop: 699.
FEATURES             Location/Qualifiers
     source          1..788
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4658071"
                     /clone_lib="NIH_MGC_61"
                     /tissue_type="embryonal carcinoma"
                     /lab_host="DH10B (T1 phage resistant)"
                     /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site 1:
                     SfiI (ggccgcctcggcc); Site 2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG dT(30)BN 3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT      204 a    178 c    176 g    228 t      2 others
ORIGIN

  Query Match             24.7%;  Score 510;  DB 10;  Length 788;
  Best Local Similarity   99.7%;  Pred. No. 8.6e-253;
  Matches  610;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        972 cggctcgaccagatggtgcatcctgtagccgagcgcctggacatcctgatgtctttggtt 1031
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 CGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCGCCTGGACATCCTGATGTCTTTGGTT 61

Qy       1032 ttgtcctacatgaaggatgtctgctatgtagatggtaaggttgataacggcaaaacaaag 1091
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 TTGTCCTACATGAAGGATGTCTGCTATGTAGATGGTAAGGTTGATAACGGCAAAACAAAG 121

Qy       1092 gatctatatcgcgacctgataaacatctttgacaaactcctgttgcccacccatgcctcc 1151
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 GATCTATATCGCGACCTGATAAACATCTTTGACAAACTCCTGTTGCCCACCCATGCCTCC 181

Qy       1152 tgccatgtacagtttttcatgttttacctctgtagtttcaaattgggattcgcagaggca 1211
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 TGCCATGTACAGTTTTTCATGTTTTACCTCTGTAGTTTCAAATTGGGATTCGCAGAGGCA 241

Qy       1212 tttttggaacatctctggaaaaaattgcaggacccaagtaatcctgccatcatcaggcag 1271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TTTTTGGAACATCTCTGGAAAAAATTGCAGGACCCAAGTAATCCTGCCATCATCAGGCAG 301

Qy       1272 gctgctggaaattatattggaagcttttggcaagagctaaatttattcctcttattact 1331
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        302 GCTGCTGGAAATTATATTGGAAGCTTTNGGCAAGAGCTAAATTTATTCCTCTTATTACT 361

Qy       1332 gtaaaatcatgcctagatcttttggttaactggctgcacatataccttaataaccaggat 1391
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 GTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCAGGAT 421

Qy       1392 tcgggaacaaaggcattctgcgatgttgctctccatggaccattttactcagcctgccaa 1451
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 GCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAA 481

Qy       1452 gctgtgttctacacctttgtttttagacacaagcagcttttgagcggaaacctgaaagaa 1511
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 GCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAA 541

Qy       1512 ggtttgcagtatcttcagagtctgaattttgagcggatagtgatgagccagctaaatccc 1571
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 GGTTNGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAATCCC 601

Qy       1572 ctgaagatttgc 1583
              ||||||||||||
Db        602 CTGAAGATTTGC 613
```

```
RESULT   3
AA213789
LOCUS       AA213789                 535 bp    mRNA    linear   EST 13-AUG-1997
DEFINITION  zr91a11.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:683036 3',
            mRNA sequence.
ACCESSION   AA213789
VERSION     AA213789.1  GI:1812416
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 535)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: polyT not found
            Insert Length: 1076   Std Error: 0.00
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 457.
FEATURES             Location/Qualifiers
     source          1. .535
                     /organism="Homo sapiens"
                     /db_xref="GDB:5586356"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:683036"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      150 a    102 c    121 g    161 t      1 others
ORIGIN


  Query Match             24.2%;  Score 500;  DB 9;  Length 535;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-247;
  Matches  500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     113 cgcttcgtcctctgcagttaagaagctgggcgcgtcgaggactgggatttcaaatatgcg 172
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7 CGCTTCGTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATATGCG 66

Qy     173 tgcattagagaatgacttttttcaattctccccccaagaaaaactgttcggtttggtggaac 232
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67 TGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTTCGGTTTGGTGGAAC 126

Qy     233 tgtgacagaagtcttgctgaagtacaaaaagggtgaaacaaatgacttttgagttgttgaa 292
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127 TGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAA 186

Qy     293 gaaccagctgttagatccagacataaaggatgaccagatcatcaactggctgctagaatt 352
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     187 GAACCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAGAATT 246

Qy     353 ccgttcttctatcatgtacttgacaaaaagacttgaagcaacttatcagtattatattaaa 412
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     247 CCGTTCTTCTATCATGTACTTGACAAAAGACTTGAGCAACTTATCAGTATTATATTAAG 306

Qy     413 attgccttggttgaatagaagtcaaacagtagtggaagagtatttggcttttcttgattaa 472
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     307 ATTGCCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTGATTAA 366

Qy     473 tcttgtatcagcacagactgttttcctcagaccgtgtctcagcatgattgcttcccattt 532
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     367 TCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTT 426

Qy     533 tgtgcctccccgagtgatcattaaggaaggcgatgtagatgtttcagattctgatgatga 592
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     427 TGTGCCTCCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGATGATGA 486

Qy     593 agatgataaatcttcctgcaa 612
           ||||||||||||||||||||
Db     487 AGATGATAAATCTTCCTGCAA 506


RESULT   4
BG428305
LOCUS       BG428305                 1037 bp    mRNA    linear   EST 14 MAR 2001
DEFINITION  602498983F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4612605 5',
            mRNA sequence.
ACCESSION   BG428305
VERSION     BG428305.1  GI:13334811
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1037)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1361  row: e  column: 22
            High quality sequence stop: 637.
FEATURES             Location/Qualifiers
     source          1. .1037
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4612605"
                     /clone_lib="NIH_MGC_75"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcgagcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(40)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T).  Average insert size 1.65
                     kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      292 a    218 c    264 g    264 t
ORIGIN


  Query Match             24.0%;  Score 497;  DB 10;  Length 1037;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-246;
  Matches  497;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        49 ggtcggttagttcggcccaatggcggcaccgctgcttcacacggtttgccgggagatg 108
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2 GGTCGGTTAGTTCGGCCCAATGGCGGCACCGCTGCTTCACACGGTTTGCCGGGAGATG 61

Qy       109 cggccgcttcgtcctctgcagttaagaagctgggcgcgtcgaggactgggatttcaaata 168
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62 CGGCCGCTTCGTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATA 121

Qy       169 tgcgtgcattagagaatgactttttcaattctccccaagaaaaactgttcggtttggtg 228
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       122 TGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCAAGAAAAACTGTTCGGTTTGGTG 181

Qy       229 gaactgtgacagaagtcttgctgaagtacaaaaagggtgaaacaaatgactttgagttgt 288
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       182 GAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTGT 241

Qy       289 tgaagaaccagctgttagatccagacataaaggatgaccagatcatcaactggctgctag 348
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       242 TGAAGAACCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAG 301

Qy       349 aattccgttcttctatcatgtacttgacaaaagactttgagcaacttatcagtattatat 408
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       302 AATTCCGTTCTTCTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATAT 361

Qy       409 taagattgccttggttgaatagaagtcaaacagtagtggaagagtatttggcttttcttg 468
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       362 TAAGATTGCCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTG 421

Qy       469 gtaatcttgtatcagcacagactgtttttcctcagaccgtgtctcagcatgattgcttccc 528
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       422 GTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTCCC 481

Qy       529 attttgtgcctccccga 545
             |||||||||||||||||
Db       482 ATTTTGTGCCTCCCCGA 498
```

```
RESULT      5
AW408066
LOCUS       AW408066                 563 bp    mRNA    linear   EST 16-FEB-2000
DEFINITION  UI-HF-BM0-adu-b-10-0-UI.r1 NIH_MGC_38 Homo sapiens cDNA clone
            IMAGE:3062754 5', mRNA sequence.
ACCESSION   AW408066
VERSION     AW408066.1  GI:6927123
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 563)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Eco RI site shown at the beginning of the sequence.
             Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: M.B. Soares Lab
             cDNA Library Arrayed by: M.B. Soares Lab
             DNA Sequencing by: M.B. Soares Lab
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1..563
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3062754"
                     /clone_lib="NIH_MGC_38"
                     /tissue_type="lymph"
                     /cell_type="germinal center B cells"
                     /cell_line="MGC85"
                     /lab_host="DH10B (TI)"
                     /note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
                     Constructed from size fractionated cytoplasmic mRNA
                     (2.5-3.5kb). Directionally cloned. Cells provided by Louis
                     M. Staudt, Ph.D. Library preparation by Maria de Fatima
                     Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT      151 a    122 c    119 g    171 t
ORIGIN

  Query Match             22.6%;  Score 467;  DB 9;  Length 563;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-230;
  Matches  467;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1054 gctatgtagatggtaaggttgataacggcaaaacaaaggatctatatcgcgacctgataa 1113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         8 GCTATGTAGATGGTAAGGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAA 67

Qy      1114 acatctttgacaaactcctgttgcccacccatgcctcctgccatgtacagtttttcatgt 1173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        68 ACATCTTTGACAAACTCCTGTTGCCCACCCATGCCTCCTGCCATGTACAGTTTTTCATGT 127

Qy      1174 tttacctctgtagtttcaaattgggattcgcagaggcatttttggaacatctctggaaaa 1233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       128 TTTACCTCTGTAGTTTCAAATTGGGATTCGCAGAGGCATTTTTGGAACATCTCTGGAAAA 187

Qy      1234 aattgcaggacccaagtaatcctgccatcatcaggcaggctgctggaaattatattggaa 1293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       188 AATTGCAGGACCCAAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAA 247

Qy      1294 gctttttggcaagagctaaatttattcctcttattactgtaaaatcatgcctagatcttt 1353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       248 GCTTTTTGGCAAGAGCTAAATTTATTCCTCTTATTACTGTAAAATCATGCCTAGATCTTT 307

Qy      1354 tggttaactggctgcacatataccttaataaccaggattcgggaacaaaggcattctgcc 1413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       308 TGGTTAACTGGCTGCACATATACCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCC 367

Qy      1414 atgttgctctccatggaccatttactcagcctgccaagctgtgttctacacctttgttt 1473
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       368 ATGTTGCTCTCCATGGACCATTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTT 427

Qy      1474 ttagacacaagcagcttttgagcggaaacctgaaagaaggtttgcag 1520
             |||||||||||||||||||||||||||||||||||||||||||||||
Db       428 TTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGTTTGCAG 474

RESULT      6
BI561523
LOCUS       BI561523                 733 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603256222F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5298581 5',
            mRNA sequence.
ACCESSION   BI561523
VERSION     BI561523.1  GI:15448837
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 733)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
```

```
            Plate: LLAM11755  row: 1  column: 06
            High quality sequence stop: 717.
FEATURES             Location/Qualifiers
     source          1. .733
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5298581"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI XhoI (qtcgaq
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      205 a    142 c    171 g    214 t      1 others
ORIGIN


  Query Match             22.2%;  Score 460;  DB 10;  Length 733;
  Best Local Similarity   99.6%;  Pred. No. 7.1e-227;
  Matches  560;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      55 gttagttcggcccaatggcggcaccgctgcttcacacgcgtttgccgggagatgcggccg 114
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9 GTTAGTTCGGCCCAATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGATGCGGCCG 68

Qy     115 cttcgtcctctgcagttaagaagctgggcgcgtcgaggactgggatttcaaatatgcgtg 174
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69 CTTCGTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATATGCGTG 128

Qy     175 cattagagaatgacttttttcaattctcccccaagaaaaactgttcggtttggtggaactg 234
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     129 CATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTTCGGTTTGGTGGAACTG 188

Qy     235 tgacagaagtcttgctgaagtacaaaaagggtgaaacaaatgacttgaagttgttgaaga 294
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     189 TGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGA 248

Qy     295 accagctgttagatccagacataaaaggatgaccagatcatcaactggctgctagaattcc 354
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     249 ACCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGTTGCTAGAATTCC 308

Qy     355 gttcttctatcatgtacttgacaaaagactttgagcaacttatcagtattatattaagat 414
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     309 GTTCTTCTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGAT 368

Qy     415 tgccttggttgaatagaagtcaaacagtagtggaagagtatttggctttcttggtaatc 474
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     369 TGCCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTCTTGGTAATC 428

Qy     475 ttgtatcagcacagactgttttcctcagaccgtgtctcagcatgattgcttcccatttg 534
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     429 TTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTG 488

Qy     535 tgcctccccgagtgatcattaaggaaggcgatgtagatgtttcagattctgatgatgaag 594
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     489 TGCCTTCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGATGATGAAG 548

Qy     595 atgataatcttcctgcaaattt 616
           |||||||||||||||||||||||
Db     549 ATGATAATCTTCCTGCAAATTT 570


RESULT   7
BE549643
LOCUS       BE549643                 458 bp    mRNA    linear   EST 10-AUG-2000
DEFINITION  7b40g08.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3230750 3'
            similar to TR:O75703 O75703 HYPOTHETICAL 12.4 KD PROTEIN. ;, mRNA
            sequence.
ACCESSION   BE549643
VERSION     BE549643.1  GI:9791345
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 458)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .458
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone "IMAGE:3230750"
                     /clone_lib="NCI_CGAP_Lu24"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T4D Pac (Pharmacia) with a
                     modified polylinker; Plasmid DNA from the normalized
                     library NCI_CGAP_Lu5 was prepared, and ss circles were
                     made in vitro. Following HAP purification, this DNA was
                     used as tracer in a subtractive hybridization reaction.
                     The driver was PCR-amplified cDNAs from a pool of 5,000
                     clones made from the same library (cloneIDs
                     1414920-1417991 and 1520904-1522439). Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT      133 a     82 c    100 g    143 t
ORIGIN


  Query Match             22.1%;  Score 458;  DB 10;  Length 458;
  Best Local Similarity   100.0%;  Pred. No. 7.3e-226;
  Matches  458;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     154 gtggcatttcaaatatgcgtgcattagagaatgacttttttcaattctcccccaagaaaaaa 213
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTGGGATTTCAAATATGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAAA 60

Qy     214 gtgttcggtttggtggaactgtgacagaagtcttgctgaagtacaaaaagggtgaaacaa 273
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GTGTTCGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAA 120

Qy     274 atgacttttgaagttgttgaaagaaccagctgttagatccagacataaaaggatgaccagatca 333
           |||||||||||.||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCA 180

Qy     334 tcaactggctgctagaattccgttcttctatcatgtacttgacaaaagacttttgagcaac 393
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TCAACTGGCTGCTAGAATTCCGTTCTTCTATCATGTACTTGACAAAAGACTTTGAGCAAC 240

Qy     394 ttatcagtattatattaagattgccttggttgaaatagaaagtcaaacagtagtggaagagt 453
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TTATCAGTATTATATTAAGATTGCCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGT 300

Qy     454 atttggctttcttggtaatcttgtatcagcacagactgttttcctcagaccgtgtctca 513
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ATTTGGCTTTCTTGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCA 360

Qy     514 gcatgattgcttcccatttgtgcctccccgagtgatcattaaggaaggcgatgtagatg 573
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATGATTGCTTCCCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCGATGTAGATG 420

Qy        574 tttcagattctgatgatgaagatgataatcttcctgca 611
              ||||||||||||||||||||||||||||||||||||||
Db        421 TTTCAGATTCTGATGATGAAGATGATAATCTTCCTGCA 458


RESULT    8
AV703279
LOCUS       AV703279                 671 bp    mRNA    linear   EST 09-OCT-2000
DEFINITION  AV703279 ADB Homo sapiens cDNA clone ADBCCD12 5', mRNA sequence.
ACCESSION   AV703279
VERSION     AV703279.1  GI:10720608
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 671)
  AUTHORS   Peng,Y., Song,H., Huang,Q., Huang,C., Gu,Y., Yang,Y., Gao,G., Xiao
            ,H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z.,
            Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
            ,G., Hu,R., Chen,J., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA ADB clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86 21 50801919(ex.45)
            Fax: 86 21 50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1..671
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "ADBCCD12"
                     /clone_lib "ADB"
                     /tissue_type "Adrenal gland"
                     /dev_stage "Adult"
                     /lab_host "SOLR"
                     /note "Vector: pBluescript sk( ); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      189 a    131 c    157 g    193 t      1 others
ORIGIN


  Query Match             21.7%;  Score 449;  DB 9;  Length 671;
  Best Local Similarity   99.8%;  Pred. No. 3.5e-221;
  Matches  499;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        416 gccttggttgaatagaagtcaaacagtagtggaagagtatttggcttttcttggtaatct 475
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTGGTAATCT 60

Qy        476 tgtatcagcacagactgttttcctcagaccgtgtctcagcatgattgcttcccattttgt 535
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGT 120

Qy        536 gcctccccgagtgatcattaaggaaggcgatgtagatgtttcagattctgatgatgaaga 595
              |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db        121 GCCTCCCCGAGTGATCATTAAGGAAGGCGATGTACATGTTTCAGATTCTGATGATGAAGA 180

Qy        596 tgataatcttcctgcaaatttttgacacatgtcacagagccttgcaaataatagcaagata 655
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGATAATCTTCCTGCAAATTTTGACACATGTCACAGAGCCTTGCAAATAATAGCAAGATA 240

Qy        656 tgtaccatcgacaccgtggtttctcatgccaatactggtggaaaaatttccatttgttcg 715
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTCG 300
```

```
Qy        716 aaaatcagagagaacactggaatgttacgttcataacttactaaggattagtgtatattt 775
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAATCAGAGAGAACACTGGAATGTTACGTTCATAACTTACTAAGGATTAGTGTATATTT 360

Qy        776 tccaaccttgaggcatgaaattctggagcttattattgaaaaactactcaagttggatgt 835
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCAACCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAAACTACTCAAGTTGGATGT 420

Qy        836 gaatgcatcccggcagggtattgaagatgctgaagaaacagcaactcaaacttgtggtgg 895
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GAATGCATCCCGGCAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGGTGG 480

Qy        896 gacagattccacggaaggat 915
              ||||||||||||||||||||
Db        481 GACAGATTCCACGGAAGGAT 500


RESULT    9
AW958173
LOCUS       AW958173                 564 bp    mRNA    linear   EST 01-JUN-2000
DEFINITION  EST370243 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW958173
VERSION     AW958173.1  GI:8147856
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 117
            Seq primer: Reverse.
FEATURES             Location/Qualifiers
     source          1..564
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone_lib "MAGE resequences, MAGE"
                     /note "Vector: pBluescriptSKm"
BASE COUNT      157 a    108 c    130 g    168 t      1 others
ORIGIN


  Query Match             21.5%;  Score 444;  DB 9;  Length 564;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-218;
  Matches  444;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         82 tgcttcacacgcgtttgccgggagatgcggccgcttcgtcctctgcagttaagaagctgg 141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGCTTCACACGCGTTTGCCGGGAGATGCGGCCGCTTCGTCCTCTGCAGTTAAGAAGCTGG 60

Qy        142 gcgcgtcgaggactgggatttcaaatatgcgtgcattagagaatgacttttttcaattctc 201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCGCGTCGAGGACTGGGATTTCAAATATGCGTGCATTAGAGAATGACTTTTTTCAATTCTC 120

Qy        202 ccccaagaaaaactgttcggtttggtggaactgtgacagaagtcttgctgaagtacaaaa 261
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCCAAGAAAAACTGTTCGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAA 180

Qy        262 agggtgaaacaaatgactttgagttgttgaagaaccagctgttagatccagacataaagg 321
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAAAGG 240
```

```
Qy     322 atgaccagatcatcaactggctgctagaattccgttcttctatcatgtacttgacaaaag 381
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATGTACTTGACAAAAG 300

Qy     382 actttgaagcaacttatcagtattatattaagattgccttggttgaatagaagtcaaacag 441
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ACTTTGAGCAACTTATCAGTATTATATTAAGATTGCCTTGGTTGAATAGAAGTCAAACAG 360

Qy     442 tagtggaagagtatttggctttttcttggtaatcttgtatcagcacagactgttttcctca 501
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TAGTGGAAGAGTATTTGGCTTTTCTTGGTAATCTTGTATCAGCACAGACTGTTTTCCTCA 420

Qy     502 gaccgtgtctcagcatgattgctt 525
           ||||||||||||||||||||||||
Db     421 GACCGTGTCTCAGCATGATTGCTT 444
```

```
RESULT  10
BG611364
LOCUS       BG611364                 697 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602612725F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4738063 5',
            mRNA sequence.
ACCESSION   BG611364
VERSION     BG611364.1  GI:13662735
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 697)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1607  row: a  column: 08
            High quality sequence stop: 667.
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4738063"
                     /clone_lib="NIH_MGC_60"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcgagcc); Site_2: SfiI (ggccattatggcc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT      178 a    170 c    170 g    179 t
ORIGIN
```

```
  Query Match             20.1%;  Score 415;  DB 10;  Length 697;
  Best Local Similarity   99.8%;  Pred. No. 1.5e-203;
  Matches  465;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    1603 tttttgctgcaatcacaaataagtaccagctcgtcttctgctacaccatcattgagagga 1662
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79 TTTTTGCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGA 138

Qy    1663 acaatcgccagatgctgccagtcattaggagtaccgctggaggagactcagtgcagatct 1722
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 ACAATCGCCAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCT 198

Qy    1723 gcacaaacccgctggacaccttcttccccttgatccctgtgtgctgaagaggtcaaaga 1782
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 GCACAAACCCGCTGGACACCTTCTTCCCCTTGATCCCTGTGTGCTGAAGAGGTCAAAGA 258

Qy    1783 aattcattgatcctatttatcaggtatgggaagacatgagtgctgaagaggctacaggagt 1842
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     259 AATTCATTGATCCTATTTATCAGGTGTGGGAAGACATGAGTGCTGAAGAGGCTACAGGAGT 318

Qy    1843 tcaagaaacccatgaaaaaggacataqtggaagatgaagatgatgacttttctgaaaggcg 1902
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 TCAAGAAACCCATGAAAAAGGACATAGTGGAAGATGAAGATGATGACTTTCTGAAAGGCG 378

Qy    1903 aagtgccccagaatgataccgtgattgggatcacaccaaggctcctttgacacgcatttcc 1962
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     379 AAGTGCCCCAGAATGATACCGTGATTGGGATCACACCAAGGCTCCTTTGACACGCATTTCC 438

Qy    1963 gaagtcctttcaaagtagtgtgggctccccacccgtgttgtacatgcaacccagtcccctct 2022
           ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 GAAGTCCTTCAAGTAGTGTGGGCTCCCCACCCGTGTTGTACATGCAACCCAGTCCCCTCT 498

Qy    2023 gacggcagaaatttgtgactgagatgtgacatttggggattccccat 2068
           ||||||||||||||||||||||||||||||||||||||||||||||
Db     499 GACGGCAGAAATTTGTGACTGAGATGTGACATTTGGGATTCCCCAT 544
```

```
RESULT  11
AA811628
LOCUS       AA811628                 440 bp    mRNA    linear   EST 19-FEB-1998
DEFINITION  ob74d03.s2 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1337093 3',
            mRNA sequence.
ACCESSION   AA811628
VERSION     AA811628.1  GI:2881239
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 440)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 823    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 424.
FEATURES             Location/Qualifiers
     source          1. .440
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1337093"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
```

was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3']. Double stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 130 a 82 c 95 g 133 t
ORIGIN

Query Match 18.8%; Score 389; DB 9; Length 440;
Best Local Similarity 99.8%; Pred. No. 4.2e-190;
Matches 439; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy     195 aattctccccccaaqaaaaactqttcqqttttqqtqqaactqtqacagaaqtcttqctqaaq 254
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AATTCTCCCCCAAGAAAAACTGTTCGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAG 60

Qy     255 tacaaaaaqqqtqaaacaaatqacttttqaqttqttqaaqaaccaqctqttqatccaqac 314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGAC 120

Qy     315 ataaaqqatqaccaqatcatcaactqqctqctaqaattccqttcttctatcatqtacttq 374
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ATAAAGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATGTACTTG 180

Qy     375 acaaaaqactttqaqcaacttatcaqtattatattaaqattqccttqqttqaataqaaqt 434
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCCTTGGTTGAATAGAAGT 240

Qy     435 caaacaqtaqtqaaaqaqtatttqqctttttcttqqtaatctttqtatcaqcacaqactqtt 494
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTGGTAATCTTGTATCAGCACAGACTGTT 300

Qy     495 ttcctcaqaccqtqtctcaqcatqattqcttcccatttqtqcctccccqaqtqatcatt 554
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TTCCTCAGACGGTGTCTCAGCATGATTGCTTCCCATTTGTGCCTCCCCGAGTGATCATT 360

Qy     555 aaqqaaqqcqatqtaqatqtttcaqattctqatqatqaaqatqataattcttcctqcaaat 614
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AAGGAAGGCGATGTAGATGTCTCAGATTCTGATGATGAAGATGATAATTCTTCCTGCAAAT 420

Qy     615 tttqacacatqtcacaqaqc 634
           ||||||||||||||||||||
Db     421 TTTGACACATGTCACAGAGC 440
```

RESULT 12
BE203684
LOCUS BE203684 996 bp mRNA linear EST 06-NOV-2000
DEFINITION 601866431F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4099398 5', mRNA sequence.
ACCESSION BE203684
VERSION BE203684.1 GI:11097270
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 996)
AUTHORS NIH MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM967 row: n column: 07
High quality sequence stop: 684.

FEATURES Location/Qualifiers
source 1..996
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4099398"
/clone_lib="NIH_MGC_17"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 338 a 211 c 241 g 206 t
ORIGIN

Query Match 18.6%; Score 385; DB 10; Length 996;
Best Local Similarity 99.6%; Pred. No. 5.6e-188;
Matches 485; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy    1297 cqcaqqacccaaqtaatcctqccatcatcaqqcaqqctqctqqaaattatattqqaaqct 1356
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96 TGCAGGACGCAAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCT 95

Qy    1357 ttttqqcaaqaqctaaattttattcctcttattactqtaaaatcatqcctaqatcttttqq 1416
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db      96 TTGGGCAAGAGCTAAATTTATTCCTCTTATTACTGTAAAATCATGCCTAGATCTTGTGG 155

Qy    1417 ttaactqqctqcacatatacctttaataaccaqqattcqqqaacaaaaqqcattctqcaatq 1476
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     156 TAACTGGCTGCACATATACCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATG 215

Qy    1477 ttqctctccatqqaccatttttactcaqcctqccaaqctqtqttctacaccttqttttta 1536
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     216 TTGCTCTCCATGGACCATTTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTGTTTTTA 275

Qy    1537 tacacaaqcaqctttqaqcqqaaacctqaaaqaaqctttqcaqtatcttcaqaqtctqa 1596
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     276 CACACAAGCAGCTTTGAGCGGAAACCTGAAAGAAGCTTTGCAGTATCTTCAGAGTCTGA 335

Qy    1597 atttqaqcqqatqtqatqaqccaqctaaatcccctqaaqatttqcctqcctcaqtqq 1656
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     336 ATTTTGAGCGGATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGG 395

Qy    1657 ttaacttttttqctqcaatcacaaataaqtaccaqctcqtcttctqctacaccatcattq 1716
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     396 TTAACTTTTTGCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTG 455

Qy    1717 aqaatcq 1723
           |||||||
Db     516 AGATCTG 522
```

RESULT 13
AA481295/c
LOCUS AA481295 380 bp mRNA linear EST 14-AUG-1997
DEFINITION zv82e05.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:746336 5', mRNA sequence.
ACCESSION AA481295
VERSION AA481295.1 GI:2210847

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 380)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 361.
FEATURES             Location/Qualifiers
     source          1. .380
                     /organism="Homo sapiens"
                     /db_xref="GDB:5945860"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:746336"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      105 a     90 c     76 g    109 t
ORIGIN


  Query Match             18.4%;  Score 380;  DB 9;  Length 380;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-185;
  Matches  380;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     854 tattgaaagatgctgaagaaacagcaactcaaacttgtggtgggacagattccacggaagg 913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     380 TATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGGTGGGACAGATTCCACGGAAGG 321

Qy     914 attgtttaatatggatgaaagatgaaagaaactgaacatgaaacaaagggctggtcctgaacg 973
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     320 ATTGTTTAATATGGATGAAGATGAAGAAACTGAACATGAAACAAAGGCTGGTCCTGAACG 261

Qy     974 gctcgaccagatggtgcatcctgtagccgagcgcctggacatcctgatgtctttggtttt 1033
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     260 GCTCGACCAGATGGTGCATCCTGTAGCCGAGCGCCTGGACATCCTGATGTCTTTGGTTTT 201

Qy    1034 gtcctacatgaaggatgtctgctatgtagatggtaaggttgataacggcaaaacaaagga 1093
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     200 GTCCTACATGAAGGATGTCTGCTATGTAGATGGTAAGGTTGATAACGGCAAAACAAAGGA 141

Qy    1094 tctatatcgcgacctgataaacatctttgacaaactcctgttgcccaccatgcctcctg 1153
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     140 TCTATATCGCGACCTGATAAACATCTTTGACAAACTCCTGTTGCCCACCCATGCCTCCTG 81
```

```
Qy    1154 ccatgtacagtttttcatatttacctctgtagtttcaaattgggattcggcagaggcatt 1213
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80 CCATGTACAGTTTTTCATGTTTTACCTCTGTAGTTTCAAATTGGGATTCGGCAGAGGCATT 21

Qy    1214 tttggaaacatctctggaaaa 1233
           |||||||||||||||||||||
Db      20 TTTGGAACATCTCTGGAAAA 1


RESULT  14
BE502966
LOCUS       BE502966                 490 bp    mRNA    linear   EST 04 AUG 2000
DEFINITION  hz81q07.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3214428 3'
            similar to PIR:O75703 O75703 HYPOTHETICAL 12.4 KD PROTEIN. ;, mRNA
            sequence.
ACCESSION   BE502966
VERSION     BE502966.1  GI:9705374
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 427.
FEATURES             Location/Qualifiers
     source          1. .490
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3214428"
                     /clone_lib="NCI_CGAP_Lu24"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Plasmid DNA from the normalized
                     library NCI_CGAP_Lu5 was prepared, and ss circles were
                     made in vitro. Following HAP purification, this DNA was
                     used as tracer in a subtractive hybridization reaction.
                     The driver was PCR-amplified cDNAs from a pool of 5,000
                     clones made from the same library (cloneIDs
                     1414920-1417991 and 1520904-1522439). Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT      146 a     90 c    101 g    151 t      2 others
ORIGIN


  Query Match             18.1%;  Score 375;  DB 10;  Length 490;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-183;
  Matches  375;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     105 aatttctccccccaaagaaaaaactgtttcggtttggtggaaactgtgacagaaggtctttgctgaag 254
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AATTCTCCCCCAAGAAAAACTGTTCGGTTTGGTGGAACTGTGACAGAAGGTCTTTGCTGAAG 60

Qy     255 tacaaaaaaggtgtgaaacaaatgactttgaagttgttgaagaaccagctgttagatccagac 314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGAC 120

Qy     315 ataaaggatgaccagatcatcaactggctgctagaattccgttcttctatcatgtacttg 374
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATAAAGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATGTACTTG 180

Qy    475 acaaaaqactttqaqcaacttatcaqtattatatattaaqattqccttqqttqaataqaaqt 434
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCCTTGGTTGAATAGAAGT 240

Qy    435 caaacaqtaqtqaaaaaqtatttqqctttcttqqtaatcttqtatcaqcacaqactqtt 494
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAAACAGTAGTGGAAGATATTTGGCTTTCTTGGTAATCTTGTATCAGCACAGACTGTT 300

Qy    495 ttcctcaqaccqtqtctcaqcatqattqcttccattttqtqcctcccqaqtqatcatt 554
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTCCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGTGCCTCCCCGAGTGATCATT 360

Qy    555 aaqqaaqqcqatqta 569
          |||||||||||||||
Db    361 AAGGAAGGCGATGTA 375


RESULT 15
BG167061
LOCUS       BG167061                  990 bp    mRNA    linear   EST 06-FEB-2001
DEFINITION  602344740F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4455009 5',
            mRNA sequence.
ACCESSION   BG167061
VERSION     BG167061.1  GI:12674764
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 990)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10247  row: o  column: 10
            High quality sequence stop: 665.
FEATURES             Location/Qualifiers
     source          1..990
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:4455009"
                     /clone_lib "NIH_MGC_89"
                     /tissue_type "hypernephroma, cell line"
                     /lab_host "DH10B (phage resistant)"
                     /note "Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo dT primed.
                     Average insert size 1.3 kb. Library enriched for
                     full length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      304 a    275 c    219 g    192 t
ORIGIN


  Query Match             18.0%;  Score 473;  DB 10;  Length 990;
  Best Local Similarity   99.8%;  Pred. No. 9.2e-182;
  Matches  423;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1568 tcccctqaaqatttqcctqcccctcaqtqqttaacttttttqctqccaatcacaaataaqta 1627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    107 TCGCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTGCTGCCAATCACAAATAAGTA 166

Qy   1628 ccaqctcqtctttctqctacacaccatcattqaqaqqaacaatcqccaqatqctqccaqtcat 1687
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    167 CCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAATCGCCAGATGCTGCCAGTCAT 226

Qy   1688 taqaaqtaccqctqqaqqaqactcaqtqcaqatctqcacaaaccccqctqqacaccttctt 1747
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    227 TAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACACCTTCTT 286

Qy   1748 ccccttqatccctgtgtqctqaaqaqqtcaaaqaaattcattqatcctatttatcaqqt 1807
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    287 CCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTTATCAGGT 346

Qy   1808 atqqqaaqacatqaqtqctqaaqaqctacaqqaqttcaaqaaaccccatqaaaaaqqacat 1867
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    347 GTGGGAAGACATGAGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAAGGACAT 406

Qy   1868 aqtqqaaqatqaaaqatqatqactttctqaaqqqcaaqtqcccaqaatqatacccqtqat 1927
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    407 AGTGGAAGATGAAGATGATGACTTTCTGAAAGGCCAAGTGCCCCAGAATGATACCGTGAT 466

Qy   1928 tqqaatcacaccaaqctccttqacacqcatttccqaaqtccttcaaqtaqtqtqqqctc 1987
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    467 TGGCATCACACCAAGCTCCTTTGACACGCATTTCCGAAGTCCTTCAAGTAGTGTGGGCTC 526

Qy   1988 scca 1991
          ||||
Db    527 CCCA 530


Search completed: July 30, 2002, 04:27:41
Job time: 5440 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 03:56:28 ; Search time 58.64 Seconds
(without alignments)
8662.519 Million cell updates/sec

Title: US-09-932-678-1
Perfect score: 2068
Sequence: 1 acagaggctgtggctggaag.....tgacatttgggattccccat 2068

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	19	0.9	479	4	US-08-936-165A-192	Sequence 192, App
	2	19	0.9	787	1	US-08-034-245-11	Sequence 11, Appl
	3	19	0.9	37950	4	US-09-338-907-183	Sequence 183, App
	4	19	0.9	37950	4	US-09-218-207-183	Sequence 183, App
	5	18	0.9	315	4	US-08-744-419-3	Sequence 3, Appli
	6	18	0.9	540	4	US-08-744-419-1	Sequence 1, Appli
c	7	18	0.9	708	1	US-08-078-090-24	Sequence 24, Appl
c	8	18	0.9	1065	1	US-08-249-554-1	Sequence 1, Appli
c	9	18	0.9	1065	1	US-08-249-555A-1	Sequence 1, Appli
c	10	18	0.9	1065	1	US-08-249-584-1	Sequence 1, Appli
c	11	18	0.9	1065	1	US-08-734-792-1	Sequence 1, Appli
c	12	18	0.9	1065	1	US-08-078-090-1	Sequence 1, Appli
c	13	18	0.9	1065	5	PCT-US95-03628-1	Sequence 1, Appli
c	14	18	0.9	1065	5	PCT-US95-03789-1	Sequence 1, Appli
c	15	18	0.9	1446	1	US-08-596-024-5	Sequence 5, Appli
c	16	18	0.9	1446	4	US-09-020-818-5	Sequence 5, Appli
c	17	18	0.9	1446	4	US-08-907-740-5	Sequence 5, Appli
	18	18	0.9	1791	1	US-08-245-294-7	Sequence 7, Appli
	19	18	0.9	1791	1	US-08-474-499-7	Sequence 7, Appli
	20	18	0.9	1791	1	US-08-307-279A-7	Sequence 7, Appli
	21	18	0.9	1791	5	PCT-US95-06211-7	Sequence 7, Appli
c	22	18	0.9	6873	4	US-09-131-028A-1	Sequence 1, Appli
	23	18	0.9	6873	4	US-09-131-028A-8	Sequence 8, Appli
	24	18	0.9	8430	4	US-09-131-028A-6	Sequence 6, Appli
c	25	18	0.9	8430	4	US-09-131-028A-10	Sequence 10, Appl
c	26	18	0.9	10607	1	US-08-078-090-3	Sequence 3, Appli
	27	17	0.8	27	3	US-08-468-011A-5	Sequence 5, Appli
	28	17	0.8	27	4	US-09-236-468A-5	Sequence 5, Appli
	29	17	0.8	27	5	PCT-US95-07085-5	Sequence 5, Appli
c	30	17	0.8	60	1	US-07-670-296-19	Sequence 19, Appl
c	31	17	0.8	60	1	US-08-093-781-20	Sequence 20, Appl
	32	17	0.8	159	1	US-08-485-455D-78	Sequence 78, Appl
	33	17	0.8	159	2	US-08-482-130C-78	Sequence 78, Appl
	34	17	0.8	159	2	US-08-484-211C-78	Sequence 78, Appl
	35	17	0.8	159	3	US-08-906-769-78	Sequence 78, Appl
	36	17	0.8	159	3	US-08-906-616-78	Sequence 78, Appl
	37	17	0.8	159	3	US-08-817-795-78	Sequence 78, Appl
	38	17	0.8	159	3	US-08-485-443B-78	Sequence 78, Appl
	39	17	0.8	159	3	US-08-639-075A-78	Sequence 78, Appl
	40	17	0.8	159	4	US-09-012-431-78	Sequence 78, Appl
	41	17	0.8	159	4	US-09-012-692-78	Sequence 78, Appl
	42	17	0.8	159	4	US-08-906-613-78	Sequence 78, Appl
	43	17	0.8	159	5	PCT-US95-14442A-78	Sequence 78, Appl
	44	17	0.8	553	4	US-09-227-357-94	Sequence 94, Appl
	45	17	0.8	558	2	US-08-896-365-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-936-165A-192/c
; Sequence 192, Application US/08936165A
; Patent No. 6348582
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Burnham, Martin
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Lonetto, Michael
; APPLICANT: Nicholas, Richard
; APPLICANT: Pratt, Julie
; APPLICANT: Reichard, Richard
; APPLICANT: Rosenberg, Martin
; APPLICANT: Ward, Judith
; TITLE OF INVENTION: No. 6348582e1 Prokaryotic Polynucleotides,
; TITLE OF INVENTION: Polypeptides and Their Uses
; NUMBER OF SEQUENCES: 534
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,165A
; FILING DATE: 24-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/027,032
; FILING DATE: 24-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50549
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 192:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 479 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
US 08 936 165A 192


  Query Match                 0.9%;  Score 19;  DB 4;  Length 479;
  Best Local Similarity   100.0%;  Pred. No. 7.5;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     929 tgaagatgaagaagatgaa 947
           ||||||||||||||||||| 
Db     286 TGAACATGAAGAAACTGAA 268


RESULT   2
US 08 034 245 11
; Sequence 11, Application US/08034245
; Patent No. 5504197
;  GENERAL INFORMATION:
;     APPLICANT:  Schubert, David
;     APPLICANT:  Fisher, Wolfgang H.
;     TITLE OF INVENTION:  NEUROTROPHIC GROWTH FACTOR AND METHODS
;     TITLE OF INVENTION:  OF TREATMENT
;     NUMBER OF SEQUENCES:  14
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;       STREET:  444 South Flower Street, Suite 2000
;       CITY:  Los Angeles
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/034,245
;       FILING DATE:  19930422
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/590,359
;       FILING DATE:  27-SEP-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/541,276
;       FILING DATE:  20-JUN-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Reiter, Stephen E.
;       REGISTRATION NUMBER:  31,192
;       REFERENCE/DOCKET NUMBER:  P41 9369
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  619-546-4747
;       TELEFAX:  619 546 9392
;   INFORMATION FOR SEQ ID NO:  11:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  787 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  double
;       TOPOLOGY:  both
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  49..594
;       OTHER INFORMATION:  /product= "Rat K33 PDGF associated
;       OTHER INFORMATION:  protein"
US 08 034 245 11


  Query Match                 0.9%;  Score 19;  DB 1;  Length 787;
  Best Local Similarity   100.0%;  Pred. No. 7.5;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1872 gaagatgaagatgatgact 1890
           ||||||||||||||||||| 
Db     238 GAAGATGAAGATGATGACT 256


RESULT   3
US 09-338-907 183
; Sequence 183, Application US/09338907
; Patent No. 6265546
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT: Ilya, Chumakov
;  APPLICANT: Bouquleret, Lydie
;  TITLE OF INVENTION: PROSTATE CANCER GENE
;  FILE REFERENCE: GENSET.18CP1CP
;  CURRENT APPLICATION NUMBER: US/09/338,907
;  CURRENT FILING DATE: 1999-06-23
;  EARLIER APPLICATION NUMBER: 08/996,306
;  EARLIER FILING DATE: 1997-12-22
;  EARLIER APPLICATION NUMBER: 60/099,658
;  EARLIER FILING DATE: 1998-09-09
;  EARLIER APPLICATION NUMBER: 09/218,207
;  EARLIER FILING DATE: 1998-12-22
;  NUMBER OF SEQ ID NOS: 578
;  SOFTWARE: Patent.pm
; SEQ ID NO 183
;  LENGTH: 37950
;  TYPE: DNA
;  ORGANISM: Mus musculus
;  FEATURE:
;  NAME/KEY: exon
;  LOCATION: 5259..5428
;  OTHER INFORMATION: exon2
;  FEATURE:
;  NAME/KEY: exon
;  LOCATION: 12675..12791
;  OTHER INFORMATION: exon3
;  FEATURE:
;  NAME/KEY: exon
;  LOCATION: 14621..14710
;  OTHER INFORMATION: exon4
;  FEATURE:
;  NAME/KEY: exon
;  LOCATION: 19822..19912
;  OTHER INFORMATION: exon5
;  FEATURE:
;  NAME/KEY: exon
;  LOCATION: 21789..21950
;  OTHER INFORMATION: exon6
;  FEATURE:
;  NAME/KEY: exon
;  LOCATION: 23387..23510
;  OTHER INFORMATION: exon7
;  FEATURE:
;  NAME/KEY: exon
;  LOCATION: 25520..26016
;  OTHER INFORMATION: exon8
US-09-338-907-183


  Query Match                 0.9%;  Score 19;  DB 4;  Length 37950;
  Best Local Similarity   100.0%;  Pred. No. 7.9;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1320 cctcttattactgtaaaat 1338
           ||||||||||||||||||| 
Db    1077 cctcttattactgtaaaat 1095


RESULT   4
US-09-218-207-183
; Sequence 183, Application US/09218207
; Patent No. 6346381
```

```
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT: Ilya, Chumakov
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: Prostate cancer gene
;  FILE REFERENCE: GENSET.018CP1
;  CURRENT APPLICATION NUMBER: US/09/218,207
;  CURRENT FILING DATE: 1998-12-22
;  EARLIER APPLICATION NUMBER: 08/996,306
;  EARLIER FILING DATE: 1997-12-22
;  EARLIER APPLICATION NUMBER: 60/099,658
;  EARLIER FILING DATE: 1998-09-09
;  NUMBER OF SEQ ID NOS: 578
;  SOFTWARE: Patent.pm
; SEQ ID NO 183
;   LENGTH: 37950
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 5259..5328
;   OTHER INFORMATION: exon2
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 12675..12791
;   OTHER INFORMATION: exon3
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 14621..14710
;   OTHER INFORMATION: exon4
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 19822..19912
;   OTHER INFORMATION: exon5
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 21789..21950
;   OTHER INFORMATION: exon6
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 23387..23510
;   OTHER INFORMATION: exon7
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 25520..26016
;   OTHER INFORMATION: exon8
US-09-218-207-183


  Query Match             0.9%;  Score 19;  DB 4;  Length 37950;
  Best Local Similarity   100.0%;  Pred. No. 7.9;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1320 cctcttattactgtaaaat 1338
           ||||||||||||||||||||
Db    1077 cctcttattactgtaaaat 1095


RESULT   5
US-08-744-419-3
; Sequence 3, Application US/08744419
; Patent No. 6274342
;  GENERAL INFORMATION:
;    APPLICANT:  Gutierrez-Ramos et al.
;    TITLE OF INVENTION:  Monocyte Chemotactic Protein 5 (MCP-5) Molecules      and Use
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, suite 510
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/744,419
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. Provisional
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Arnold, Beth E.
;      REGISTRATION NUMBER:  35,430
;      REFERENCE/DOCKET NUMBER:  MIQ-008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  315 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-744-419-3


  Query Match             0.9%;  Score 18;  DB 4;  Length 315;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1824 gctgaaagagctacaggag 1841
           ||||||||||||||||||
Db     138 GCTGAAGAGCTACAGGAG 155


RESULT   6
US-08-744-419-1
; Sequence 1, Application US/08744419
; Patent No. 6274342
;  GENERAL INFORMATION:
;    APPLICANT:  Gutierrez-Ramos et al.
;    TITLE OF INVENTION:  Monocyte Chemotactic Protein 5 (MCP-5) Molecules      and U
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, suite 510
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/744,419
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. Provisional
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Arnold, Beth E.
;      REGISTRATION NUMBER:  35,430
;      REFERENCE/DOCKET NUMBER:  MIQ-008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
```

```
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  540 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  56..367
US-08-744-419-1


  Query Match                 0.9%;  Score 18;  DB 4;  Length 540;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1824 gctgaagagctacaggag 1841
           ||||||||||||||||||
Db     193 GCTGAAGAGCTACAGGAG 210


RESULT   7
US-08-078-090-24/c
; Sequence 24, Application US/08078090
; Patent No. 5739407
;  GENERAL INFORMATION:
;    APPLICANT:  BERGSTROEM, SVEN
;    APPLICANT:  HERNELL, OLLE
;    APPLICANT:  LOENNERDAL, BO
;    APPLICANT:  BJALMARSSON, KARIN
;    APPLICANT:  HANSSON, LENNART
;    APPLICANT:  TOERNELL, JAN
;    APPLICANT:  STROEMQUIST, MATS
;    TITLE OF INVENTION:  HUMAN BETA-CASEIN PROCESS FOR PRODUCING
;    TITLE OF INVENTION:  IT AND USE THEREOF
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 SEVENTH STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/078,090
;      FILING DATE:  19930618
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/DK92/00246
;      FILING DATE:  19 AUG 1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/DK91/00244
;      FILING DATE:  19 AUG 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  COOPER, IVER P.
;      REGISTRATION NUMBER:  28,005
;      REFERENCE/DOCKET NUMBER:  BERGSTROM2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)628 5197
;      TELEFAX:  (202)737 3528
;      TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  708 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
US-08-078-090-24


  Query Match                 0.9%;  Score 18;  DB 1;  Length 708;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cagaagctctgaggctgaaa 19
           ||||||||||||||||||
Db     213 CAGAGGCTGTGGCTGGAA 196


RESULT   8
US-08-249-554-1/c
; Sequence 1, Application US/08249554
; Patent No. 5506209
;  GENERAL INFORMATION:
;    APPLICANT:  Mukerji, P.
;    APPLICANT:  Seo, A.
;    APPLICANT:  Anderson, S.
;    APPLICANT:  Schaller, J.
;    TITLE OF INVENTION:  Product for the Inhibition of Infection of Mammalian Cells
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lonnie R.  Drayer
;      ADDRESSEE:  Ross Products Division
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  625 Cleveland Avenue
;      CITY:  Columbus
;      STATE:  Ohio
;      COUNTRY:  United States
;      ZIP:  43215
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb storage (B)COMPUTER: Apple Macintosh
;      OPERATING SYSTEM:  MacIntosh System 7.1(D)SOFTWARE: ClarisWorks 1.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/249,554
;      FILING DATE:  26-MAY 1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:  No. 5506209 applicable
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (614) 624-3774
;      TELEFAX:  (614) 624 3074
;      TELEX:  No. 5506209e
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1065 base pairs
;      TYPE:  Nucleic acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Unknown
;    MOLECULE TYPE:  Cloned cDNA representing the product of a human
;    MOLECULE TYPE:  genomic DNA segment.
;      DESCRIPTION:  Human milk Beta-casein
;    HYPOTHETICAL:
;    ANTI-SENSE:
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:  Human
;      ORGANISM:  Homo Sapiens
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:  Adult
;      HAPLOTYPE:
;      TISSUE TYPE:  Mammary Gland
;      CELL TYPE:
;      CELL LINE:
;      ORGANELLE:
;    IMMEDIATE SOURCE:  Human Mammary Gland
;      LIBRARY:
;      CLONE:
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:
;      MAP POSITION:
;        NTIS:
```

```
;     FEATURE:
;       NAME/KEY:
;       LOCATION:
;       IDENTIFICATION METHOD:  DNA sequencing and restriction analysis
;       OTHER INFORMATION:  The encoded product of nucleotide SEQ ID NO: 1: is the human
;     PUBLICATION INFORMATION:
;       AUTHORS:  B.  Lonnerdal et al
;       TITLE:  Cloning and Sequencing of a cDNA encoding human
;       TITLE:  milk beta-casein.
;       JOURNAL:  Federation European Biochemical Society Letters
;       VOLUME:  269
;       ISSUE:
;       PAGES:  153 - 156
;       DATE:  1990
;       DOCUMENT NUMBER:
;       FILING DATE:
;       PUBLICATION DATE:
;       RELEVANT RESIDUES IN SEQ ID NO:
US-08-249-554-1


  Query Match             0.9%;  Score 18;  DB 1;  Length 1065;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cagaggctgtggctggaa 19
           ||||||||||||||||||
Db     192 CAGAGGCTGTGGCTGGAA 175


RESULT   9
US-08-249-555A-1/c
; Sequence 1, Application US/08249555A
; Patent No. 5538952
;  GENERAL INFORMATION:
;    APPLICANT:  Mukerji, P.
;    APPLICANT:  Seo, A.
;    APPLICANT:  Anderson, S.
;    APPLICANT:  Schaller, J.
;    TITLE OF INVENTION:  Inhibition of Infection of Mammalian Cells by Respiratory Sync
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lonnie R.  Drayer
;      ADDRESSEE:  ROSS Products Division
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  625 Cleveland Avenue
;      CITY:  Columbus
;      STATE:  Ohio
;      COUNTRY:  United States
;      ZIP:  43215
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb storage (B)COMPUTER: Apple Macintosh
;      OPERATING SYSTEM:  MacIntosh System 7.1(D)SOFTWARE: ClarisWorks 1.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/249,555A
;      FILING DATE:  26-MAY-1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:  No. 5538952 applicable
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (614) 624-3774
;      TELEFAX:  (614) 624-3074
;      TELEX:  No. 5538952e
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1065 base pairs
;      TYPE:  Nucleic acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Unknown
;    MOLECULE TYPE:  Cloned cDNA representing the product of a human
;    MOLECULE TYPE:  genomic DNA segment.
;      DESCRIPTION:  Human milk aeta-casein
;    HYPOTHETICAL:
;    ANTI-SENSE:
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:  Human
;      ORGANISM:  Homo sapiens
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:  Adult
;      HAPLOTYPE:
;      TISSUE TYPE:  Mammary gland
;      CELL TYPE:
;      CELL LINE:
;      ORGANELLE:
;    IMMEDIATE SOURCE:  Human Mammary Gland
;      LIBRARY:
;      CLONE:
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:
;      MAP POSITION:
;      UNITS:
;    FEATURE:
;      NAME/KEY:
;      LOCATION:
;      IDENTIFICATION METHOD:  DNA sequencing and restriction analysis
;      OTHER INFORMATION:  The encoded product of nucleotide SEQ ID NO: 1: is the hum
;    PUBLICATION INFORMATION:
;      AUTHORS:  B.  Lonnerdal et al
;      TITLE:  Cloning and sequencing of a cDNA encoding human
;      TITLE:  milk beta-casein.
;      JOURNAL:  Federation European Biochemical Society Letters
;      VOLUME:  269
;      ISSUE:
;      PAGES:  153 - 156
;      DATE:  1990
;      DOCUMENT NUMBER:
;      FILING DATE:
;      PUBLICATION DATE:
;      RELEVANT RESIDUES IN SEQ ID NO:
US-08-249-555A-1


  Query Match             0.9%;  Score 18;  DB 1;  Length 1065;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cagaggctgtggctggaa 19
           ||||||||||||||||||
Db     192 CAGAGGCTGTGGCTGGAA 175


RESULT  10
US-08-249-584-1/c
; Sequence 1, Application US/08249584
; Patent No. 5643880
;  GENERAL INFORMATION:
;    APPLICANT:  Mukerji, P.
;    APPLICANT:  Seo, A.
;    APPLICANT:  Anderson, S.
;    APPLICANT:  Harvey, L.
;    TITLE OF INVENTION:  Product for Inhibition of Attachment of
;    TITLE OF INVENTION:  H.  Influenzae to Human Cells.
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lonnie R.  Drayer
;      ADDRESSEE:  ROSS Products Division
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  625 Cleveland Avenue
;      CITY:  Columbus
;      STATE:  Ohio
;      COUNTRY:  United States
;      ZIP:  43215
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb storage
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  MacIntosh System 7.1
```

```
;       SOFTWARE:  ClarisWorks 1.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/249,584
;       FILING DATE:  26-MAY 1994
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:  No. 5643880 applicable
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (614) 624-4774
;       TELEFAX:  (614) 624 3074
;       TELEX:  No. 5643880e
;   INFORMATION FOR SEQ ID No:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1065 base pairs
;       TYPE:  Nucleic acid
;       STRANDEDNESS:  Single
;       TOPOLOGY:  Unknown
;     MOLECULE TYPE:  Cloned cDNA representing the product of a human
;     MOLECULE TYPE:  genomic DNA segment.
;       DESCRIPTION:  Human milk 'eta-casein
;     HYPOTHETICAL:
;     ANTI SENSE:
;     FRAGMENT TYPE:
;     ORIGINAL SOURCE:  Human
;       ORGANISM:  Homo sapiens
;       STRAIN:
;       INDIVIDUAL ISOLATE:
;       DEVELOPMENTAL STAGE:  Adult
;       HAPLOTYPE:
;       TISSUE TYPE:  Mammary gland
;       CELL TYPE:
;       CELL LINE:
;       ORGANELLE:
;     IMMEDIATE SOURCE:  Human Mammary Gland
;       LIBRARY:
;       CLONE:
;     POSITION IN GENOME:
;       CHROMOSOME/SEGMENT:
;       MAP POSITION:
;       UNITS:
;     FEATURE:
;       NAME/KEY:
;       LOCATION:
;       IDENTIFICATION METHOD:  DNA sequencing and restriction
;       IDENTIFICATION METHOD:  analysis
;       OTHER INFORMATION:  The encoded product of nucleotide
;       OTHER INFORMATION:  SEQ ID NO:1: is the human milk protein,
;       OTHER INFORMATION:  ' casein.
;     PUBLICATION INFORMATION:
;       AUTHORS:  B. Lonnerdal et al
;       TITLE:  Cloning and sequencing of a cDNA encoding human
;       TITLE:  milk beta casein.
;       JOURNAL:  Federation European Biochemical Society Letters
;       VOLUME:  269
;       ISSUE:
;       PAGES:  154  156
;       DATE:  1990
;       DOCUMENT NUMBER:
;       FILING DATE:
;       PUBLICATION DATE:
;       RELEVANT RESIDUES IN SEQ ID NO:
US-08-249-584-1


  Query Match              0.9%;  Score 18;  DB 1;  Length 1065;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cagaacctgagcctggaa 19
           ||||||||||||||||||
Db     192 CAGAGGCTGTGGCTGGAA 175


RESULT  11
```

```
US-08-734-792-1/c
; Sequence 1, Application US/08734792
; Patent No. 5707968
;  GENERAL INFORMATION:
;    APPLICANT:  Mukerji, P.
;    APPLICANT:  Seo, A.
;    APPLICANT:  Anderson, S.
;    APPLICANT:  Harvey, L.
;    TITLE OF INVENTION:  Inhibition of Attachment of H. Influenzae
;    TITLE OF INVENTION:  to Human Cells
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lonnie R. Drayer
;      ADDRESSEE:  ROSS Products Division
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  625 Cleveland Avenue
;      CITY:  Columbus
;      STATE:  Ohio
;      COUNTRY:  United States
;      ZIP:  43215
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb storage
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh System 7.1
;      SOFTWARE:  ClarisWorks 1.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/734,792
;      FILING DATE:  23-OCT-1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/249,556
;      FILING DATE:  26-MAY-1994
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (614) 624-4774
;      TELEFAX:  (614) 624-4074
;      TELEX:  No. 5707968e
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1065 base pairs
;      TYPE:  Nucleic acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Unknown
;    MOLECULE TYPE:  Cloned cDNA representing the product of a human
;    MOLECULE TYPE:  genomic DNA segment.
;      DESCRIPTION:  Human milk 'eta casein
;    HYPOTHETICAL:
;    ANTI-SENSE:
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:  Human
;      ORGANISM:  Homo sapiens
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:  Adult
;      HAPLOTYPE:
;      TISSUE TYPE:  Mammary gland
;      CELL TYPE:
;      CELL LINE:
;      ORGANELLE:
;    IMMEDIATE SOURCE:  Human Mammary Gland
;      LIBRARY:
;      CLONE:
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:
;      MAP POSITION:
;      UNITS:
;    FEATURE:
;      NAME/KEY:
;      LOCATION:
;      IDENTIFICATION METHOD:  DNA sequencing and restriction
;      IDENTIFICATION METHOD:  analysis
;      OTHER INFORMATION:  The encoded product of nucleotide
;      OTHER INFORMATION:  SEQ ID NO:1: is the human milk protein, '-casein.
;    PUBLICATION INFORMATION:
```

```
;       AUTHORS:  B.  Lonnerdal et al
;       TITLE:  Cloning and sequencing of a cDNA encoding human
;       TITLE:  milk beta-casein
;       JOURNAL:  Federation European Biochemical Society Letters
;       VOLUME:  269
;       ISSUE:
;       PAGES:  153 - 156
;       DATE:  1990
;       DOCUMENT NUMBER:
;       FILING DATE:
;       PUBLICATION DATE:
;       RELEVANT RESIDUES IN SEQ ID NO:
US-08-734-792-1


  Query Match             0.9%;  Score 18;  DB 1;  Length 1065;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cagaggctgtggctggaa 19
           ||||||||||||||||||
Db     192 CAGAGGCTGTGGCTGGAA 175


RESULT  12
US-08-078-090-1/c
; Sequence 1, Application US/08078090
; Patent No. 5739407
;  GENERAL INFORMATION:
;     APPLICANT:  BERGSTROEM, SVEN
;     APPLICANT:  HERNELL, OLLE
;     APPLICANT:  LOENNERDAL, BO
;     APPLICANT:  HJALMARSSON, KARIN
;     APPLICANT:  HANSSON, LENNART
;     APPLICANT:  TOERNELL, JAN
;     APPLICANT:  STROEMQUIST, MATS
;     TITLE OF INVENTION:  HUMAN BETA-CASEIN PROCESS FOR PRODUCING
;     TITLE OF INVENTION:  IT AND USE THEREOF
;     NUMBER OF SEQUENCES:  24
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  BROWDY AND NEIMARK
;       STREET:  419 SEVENTH STREET, N.W.
;       CITY:  WASHINGTON
;       STATE:  D.C.
;       COUNTRY:  USA
;       ZIP:  20004
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/078,090
;       FILING DATE:  19930618
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  WO PCT/DK92/00236
;       FILING DATE:  19-AUG-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  WO PCT/DK91/00233
;       FILING DATE:  19-AUG-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  COOPER, IVER P.
;       REGISTRATION NUMBER:  28,005
;       REFERENCE/DOCKET NUMBER:  BERGSTROM2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202)628-5197
;       TELEFAX:  (202)737-3528
;       TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  1065 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  4..48
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  49..681
US-08-078-090-1


  Query Match             0.9%;  Score 18;  DB 1;  Length 1065;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cagaggctgtggctggaa 19
           ||||||||||||||||||
Db     192 CAGAGGCTGTGGCTGGAA 175


RESULT  13
PCT-US95-03628-1/c
; Sequence 1, Application PC/TUS9503628
;  GENERAL INFORMATION:
;     APPLICANT:  Abbott Laboratories
;     TITLE OF INVENTION:  Inhibition of Infection of Mammalian Cells by
;     TITLE OF INVENTION:  Respiratory Syncytial Virus
;     NUMBER OF SEQUENCES:  5
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lonnie R. Drayer
;       ADDRESSEE:  ROSS Products Division
;       ADDRESSEE:  Abbott Laboratories
;       STREET:  625 Cleveland Avenue
;       CITY:  Columbus
;       STATE:  Ohio
;       COUNTRY:  United States of America
;       ZIP:  43215
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 5.25 inch, 1.44 Mb storage
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  MS-DOS Version 6.21
;       SOFTWARE:  WordPerfect Version 6.0a
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/03628
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/249,554
;       FILING DATE:  26-MAY-1994
;       APPLICATION NUMBER:  US 08/249,555
;       FILING DATE:  26-MAY-1994
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (614) 624-3774
;       TELEFAX:  (614) 624-3074
;       TELEX:  None
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  1065 base pairs
;       TYPE:  Nucleic acid
;       STRANDEDNESS:  Single
;       TOPOLOGY:  Unknown
;     MOLECULE TYPE:  Cloned cDNA representing the product of a
;     MOLECULE TYPE:  human genomic DNA segment
;       DESCRIPTION:  Human milk  eta-casein
;     HYPOTHETICAL:
;     ANTI-SENSE:
;     FRAGMENT TYPE:
;     ORIGINAL SOURCE:  Human
;       ORGANISM:  Homo sapiens
;       STRAIN:
;       INDIVIDUAL ISOLATE:
;       DEVELOPMENTAL STAGE:  Adult
```

```
;       HAPLOTYPE:
;       TISSUE TYPE:  Mammary gland
;       CELL TYPE:
;       CELL LINE:
;       ORGANELLE:
;     IMMEDIATE SOURCE:  Human Mammary Gland
;       LIBRARY:
;       CLONE:
;     POSITION IN GENOME:
;       CHROMOSOME/SEGMENT:
;       MAP POSITION:
;       UNITS:
;     FEATURE:
;       NAME/KEY:
;       LOCATION:
;       IDENTIFICATION METHOD:  DNA sequencing and restriction analysis
;       OTHER INFORMATION:  The encoded product of nucleotide SEQ ID
;       OTHER INFORMATION:  NO: 1 is the human milk protein,  -casein.
;     PUBLICATION INFORMATION:
;       AUTHORS:  B. Lonnerdal, et al.
;       TITLE:  Cloning and sequencing of a cDNA encoding human milk beta-
;       TITLE:  casein.
;       JOURNAL:  Federation European Biochemical Society Letters
;       VOLUME:  269
;       ISSUE:
;       PAGES:  153 - 156
;       DATE:  1990
;       DOCUMENT NUMBER:
;       FILING DATE:
;       PUBLICATION DATE:
;       RELEVANT RESIDUES IN SEQ ID NO:
PCT-US95-03628-1


  Query Match              0.9%;  Score 18;  DB 5;  Length 1065;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cagaggctgtggctggaa 19
           ||||||||||||||||||
Db     192 CAGAGGCTGTGGCTGGAA 175


RESULT  14
PCT-US95-03789-1/c
; Sequence 1, Application PC/TUS9503789
;  GENERAL INFORMATION:
;     APPLICANT:  Abbott Laboratories
;     TITLE OF INVENTION:  Inhibition of Attachment of H. Influenzae to Human
;     TITLE OF INVENTION:  Cells
;     NUMBER OF SEQUENCES:  5
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Lonnie R. Drayer
;       ADDRESSEE:  ROSS Products Division
;       ADDRESSEE:  Abbott Laboratories
;       STREET:  625 Cleveland Avenue
;       CITY:  Columbus
;       STATE:  Ohio
;       COUNTRY:  United States of America
;       ZIP:  43215
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 Mb storage
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  MS DOS Version 6.21
;       SOFTWARE:  WordPerfect Version 6.0a
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/03789
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/249,556
;       FILING DATE:  26-MAY-1994
;       APPLICATION NUMBER:  US 08/249,584
;       FILING DATE:  26-MAY-1994
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (614) 624-3774
;       TELEFAX:  (614) 624-3074
;       TELEX:  None
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1065 base pairs
;       TYPE:  Nucleic acid
;       STRANDEDNESS:  Single
;       TOPOLOGY:  Unknown
;     MOLECULE TYPE:  Cloned cDNA representing the product of a
;     MOLECULE TYPE:  human genomic DNA segment
;       DESCRIPTION:  Human milk  eta-casein
;     HYPOTHETICAL:
;     ANTI SENSE:
;     FRAGMENT TYPE:
;     ORIGINAL SOURCE:  Human
;       ORGANISM:  Homo sapiens
;       STRAIN:
;       INDIVIDUAL ISOLATE:
;       DEVELOPMENTAL STAGE:  Adult
;       HAPLOTYPE:
;       TISSUE TYPE:  Mammary gland
;       CELL TYPE:
;       CELL LINE:
;       ORGANELLE:
;     IMMEDIATE SOURCE:  Human Mammary Gland
;       LIBRARY:
;       CLONE:
;     POSITION IN GENOME:
;       CHROMOSOME/SEGMENT:
;       MAP POSITION:
;       UNITS:
;     FEATURE:
;       NAME/KEY:
;       LOCATION:
;       IDENTIFICATION METHOD:  DNA sequencing and restriction analysis
;       OTHER INFORMATION:  The encoded product of nucleotide SEQ ID
;       OTHER INFORMATION:  NO: 1 is the human milk protein,  -casein.
;     PUBLICATION INFORMATION:
;       AUTHORS:  B. Lonnerdal, et al.
;       TITLE:  Cloning and sequencing of a cDNA encoding human milk beta
;       TITLE:  casein.
;       JOURNAL:  Federation European Biochemical Society Letters
;       VOLUME:  269
;       ISSUE:
;       PAGES:  153 - 156
;       DATE:  1990
;       DOCUMENT NUMBER:
;       FILING DATE:
;       PUBLICATION DATE:
;       RELEVANT RESIDUES IN SEQ ID NO:
PCT-US95-03789-1


  Query Match              0.9%;  Score 18;  DB 5;  Length 1065;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cagaggctgtggctggaa 19
           ||||||||||||||||||
Db     192 CAGAGGCTGTGGCTGGAA 175


RESULT  15
US-08-596-024-5/c
; Sequence 5, Application US/08596024
; Patent No. 5716837
;  GENERAL INFORMATION:
;     APPLICANT:  Barry, Gerard F.
;     APPLICANT:  DeWeerd, Jan W.
;     APPLICANT:  Kishore, Ganesh M.
```

```
;    APPLICANT:  Weldon, Marcia L.
;    TITLE OF INVENTION:  Expression of Sucrose Phosphorylase in
;    TITLE OF INVENTION:  Plants
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Charles E. Cohen, Monsanto Company, BB4F
;      STREET:  700 Chesterfield Parkway No. 5716837th
;      CITY:  St. Louis
;      STATE:  Missouri
;      COUNTRY:  USA
;      ZIP:  63198
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/596,024
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/386,860
;      FILING DATE:  10-FEB-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cohen, Charles E.
;      REGISTRATION NUMBER:  34,565
;      REFERENCE/DOCKET NUMBER:  38-21(13567)A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (314)537-6224
;      TELEFAX:  (314)537-6047
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1446 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-596-024-5


  Query Match             0.9%;  Score 18;  DB 1;  Length 1446;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     459 gcttttcttggtaatctt 476
           ||||||||||||||||||
Db     304 GCTTTTCTTGGTAATCTT 287



Search completed: July 30, 2002, 05:14:10
Job time: 4662 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 03:59:49 ; Search time 243.92 Seconds
(without alignments)
14556.326 Million cell updates/sec

Title: US-09-932-678-1
Perfect score: 2068
Sequence: 1 acagaggctgtggctggaag.....tgacatttgggattccccat 2068

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1869	90.4	2040	21	AAA98384	Human RNA polymera
2	920	44.5	1418	21	AAC76937	Human ORFX ORF2492
3	842	40.7	1498	22	AAS44584	Human full-length
4	691	33.4	1423	22	AAS01562	Human secretory mo
5	292	14.1	341	20	AAV90248	EST clone D1349.
6	269	13.0	3169	23	AAS87641	DNA encoding novel
7	232	11.2	1461	22	ABA83038	Human transcriptio
8	230	11.1	2103	21	AAC77798	Human cancer assoc
9	229	11.1	2991	23	AAS73660	DNA encoding novel
10	229	11.1	2991	23	AAS91992	DNA encoding novel
11	223	10.8	776	23	AAS92253	DNA encoding novel
12	215	10.4	3175	23	AAS92254	DNA encoding novel
13	208	10.1	34269	22	AAK68677	Human immune/haema
14	208	10.1	34269	22	AAK85168	Human immune/haema
15	198	9.6	2410	23	AAS92255	DNA encoding novel
16	196	9.5	437	20	AAX51663	Human secreted pro
17	190	9.2	396	23	AAS80954	DNA encoding novel
18	190	9.2	396	23	AAS87635	DNA encoding novel
19	190	9.2	396	23	AAS91986	DNA encoding novel
20	190	9.2	396	23	AAS92246	DNA encoding novel
21	187	9.0	17803	22	AAK68676	Human immune/haema
22	182	8.8	498	23	AAS92247	DNA encoding novel
23	180	8.7	349	21	AAC77533	Human ORFX ORF3088
24	176	8.5	9259	22	AAK68674	Human immune/haema
25	176	8.5	23815	22	AAK68678	Human immune/haema
26	176	8.5	23815	22	AAK85169	Human immune/haema
c 27	175	8.5	585	23	AAS92248	DNA encoding novel
28	174	8.4	1017	23	AAS87636	DNA encoding novel
c 29	139	6.7	655	23	AAS80955	DNA encoding novel
30	139	6.7	1260	23	AAS87643	DNA encoding novel
31	133	6.4	627	23	AAS87640	DNA encoding novel
32	117	5.7	213	21	AAC19091	Human secreted pro
33	110	5.3	3410	22	AAK85167	Human immune/haema
34	99	4.8	560	23	AAS91991	DNA encoding novel
c 35	91	4.4	1840	22	AAK94394	Human full-length
36	80	3.9	405	22	AAK56711	Human immune/haema
37	76	3.7	414	22	ABA46215	Human breast cell
38	76	3.7	414	22	ABA56760	Human foetal liver
39	76	3.7	414	22	ABA26382	Probe #4848 for ge
40	76	3.7	414	22	AAK04888	Human brain expres
41	76	3.7	414	22	AAK30411	Human bone marrow
42	76	3.7	414	22	AAI15019	Probe #4952 for ge
43	76	3.7	414	22	AAI36361	Probe #5047 used t
44	76	3.7	414	22	AAI04785	Probe #4776 used t
45	75	3.6	95	22	ABA51325	Human breast cell

ALIGNMENTS

RESULT 1
AAA98384
ID AAA98384 standard; cDNA; 2040 BP.
XX
AC AAA98384;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human RNA polymerase I transcription factor TIF-1A cDNA.
XX
KW RNA polymerase I transcription factor TIF-1A; antitumor; treatment;
KW antiproliferative; cell proliferation; cancer; tissue regeneration; ss.
XX
OS Homo sapiens.
XX
PN WO200055316-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-DE00767.
XX
PR 17-MAR-1999; 99DE-1011992.
XX
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI Grummt I, Vingron M;
XX
DR WPI; 2000-587527/55.
DR P-PSDB; AAB10936.
XX
PT New DNA encoding the transcription factor TIF-1A, useful for preventing
PT or treating diseases associated with abnormal cell proliferation.

```
PI   particularly tumors
XX
PS   Claim 1; Fig 2; 38pp; German.
XX
CC   This invention describes a novel DNA sequence (I) that encodes the RNA
CC   polymerase I transcription factor TIF-IA which has antitumor,
CC   antiproliferative and proliferation-inducing activity. The invention also
CC   describes (1) DNA (Ia) encoding a protein (II) with the biological
CC   activity of TIF-IA; (2) a ribozyme (R) corresponding to (I) or (Ia) and
CC   able to bind specifically to, and cleave, its transcribed RNA so as to
CC   reduce or inhibit synthesis of the corresponding protein; (3) an
CC   antisense RNA (AS) with binding properties similar to R; (4) an
CC   expression vector that contains (I), (Ia) or sequences that encode R or
CC   AS; (5) host cells containing the vectors of (4); (6) TIF-IA or (II)
CC   encoded by (I) or (Ia); (7) preparation of TIF-IA or (II) by culturing
CC   cells of (6); (8) ligands that bind to TIF IA or (II); (9) antagonists
CC   that weaken or block the activity of TIF-IA or (II); (10) a diagnostic
CC   method for detecting abnormal TIF-IA expression; and (11) kit for
CC   carrying out the method in (10). (I), and similar sequences that encode
CC   proteins with equivalent activity, expression vectors containing them, or
CC   their expression products are used to treat or prevent disorders
CC   associated with reduced cellular proliferation, to stimulate cellular
CC   proliferation, and to promote tissue regeneration, e.g. after injury or
CC   radiation therapy. Ribozymes, antisense sequences directed against (I),
CC   also ligands and antagonists of TIF-IA, are used to treat or prevent
CC   disorders associated with excessive cellular proliferation and to inhibit
CC   proliferation, especially in treatment of cancers. (I) and specific
CC   ligands for TIF IA (particularly antibodies (Ab)) are also useful for
CC   diagnosis of altered TIF-IA expression by (in)direct determination of the
CC   concentration, length and/or sequence of TIF-IA or its mRNA, e.g. for
CC   detecting mutations. Ab can also be used for immunoprecipitation of
CC   TIF-IA and for isolation of related sequences from cDNA expression
CC   libraries. (I) allows recombinant production of TIF-IA in sufficient
CC   quantities for therapeutic use.
XX
SQ   Sequence 2040 BP; 563 A; 433 C; 474 G; 568 T; 2 other;

  Query Match             90.4%;  Score 1869;  DB 21;  Length 2040;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2019;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         47 gcggtcgcgttagttcggcccaatggcggcaccgctgcttcacacgcgtttgccgggaga 106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         18 gcggtcgcgttagttcggcccaatggcggcaccgctgcttcacacgcgtttgccgggaga 77

Qy        107 tgcggccgcttcgtcctctgcagttaagaagctgggcgcgtcgaggactgggatttcaaa 166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 tgcggccgcttcgtcctctgcagttaagaagctgggcgcgtcgaggactgggatttcaaa 137

Qy        167 tatgcgtgcattagagaatgacttttcaattctcccccaagaaaaactgttcggtttgg 226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 tatgcgtgcattagagaatgacttttcaattctcccccaagaaaaactgttcggtttgg 197

Qy        227 tggaactgtgacagaagtcttgctgaagtacaaaaagggtgaaacaaatgactttgagtt 286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 tggaactgtgacagaagtcttgctgaagtacaaaaagggtgaaacaaatgactttgagtt 257

Qy        287 gttgaagaaccagctgttagatccagacataaaggatgaccagatcatcaactggctgct 346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 gttgaagaaccagctgttagatccagacataaaggatgaccagatcatcaactggctgct 317

Qy        347 agaattccgttcttctatcatgtacttgacaaaagactttgagcaacttatcagtattat 406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 agaattccgttcttctatcatgtacttgacaaaagactttgagcaacttatcagtattat 377

Qy        407 attaagattgccttggttgaatagaagtcaaacagtagtggaagagtatttggcttttct 466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 attaagattgccttggttgaatagaagtcaaacagtagtggaagagtatttggcttttct 437

Qy        467 tggtaatcttgtatcagcacagactgttttcctcagaccgtgtctcagcatgattgcttc 526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 tggtaatcttgtatcagcacagactgttttcctcagaccgtgtctcagcatgattgcttc 497

Qy        527 ccattttgtgcctccccgagtgatcattaaggaaggcgatgtagatgtttcagattctga 586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 ccattttgtgcctccccgagtgatcattaaggaaggcgatgtagatgtttcagattctga 557

Qy        587 tgatgaagatgataatcttcctgcaaatttttgacacatgtcacagagccttgcaaataat 646
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 tgatgaagatgataatcttcctgcaaatttttgacacatgtcacagagccttgcaaataat 617

Qy        647 agcaagatatgtaccatcgacaccgtggtttctcatgccaatactggtggaaaaatttcc 706
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 agcaagatatgtaccatcgacaccgtggtttctcatgccaatactggtggaaaaatttcc 677

Qy        707 atttgttcgaaaatcagagagaacactggaatgttacgttcataacttactaaggattag 766
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        678 atttgttcgaaaatcagagagaacactggaatgttacgttcatnacttactaaggattag 737

Qy        767 tgtatattttccaaccttgaggcatgaaattctggagcttattattgaaaaactactcaa 826
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        738 tgtatattttccaaccttgaggcatgaaattctggagcttattattgaaaaactactcaa 797

Qy        827 gttggatgtgaatgcatcccggcagggtattgaagatgctgaagaaacagcaactcaaac 886
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        798 gttggatgtgaatgcatcccggcagggtattgaagatgctgaagaaacagcaactcaaac 857

Qy        887 ttgtggtgggacagattccacggaaggattgtttaatatggatgaagatgaagaaaactga 946
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        858 ttgtggtgggacagattccacggaaggattgtttaatatggatgaagatgaagaaaactga 917

Qy        947 acatgaaacaaaggctggtcctgaacggctcgaccagatggtgcatcctgtagccgagcg 1006
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 acatgaaacaaaggctggtcctgaacggctcgaccagatggtgcatcctgtagccgagcg 977

Qy       1007 cctggacatcctgatgtctttggtttttgtcctacatgaaggatgtctgctatgtagatgg 1066
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 cctggacatcctgatgtctttggtttttgtcctacatgaaggatgtctgctatgtagatgg 1037

Qy       1067 taaggttgataacggcaaaacaaaggatctatatcgcgacctgataaacatctttgacaa 1126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 taaggttgataacggcaaaacaaaggatctatatcgcgacctgataaacatctttgacaa 1097

Qy       1127 actcctgttgccccacccatgcctcctgccatgtacagttttcatgttttacctctgtag 1186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 actcctgttgccccacccatgcctcctgccatgtacagttttcatgttttacctctgtag 1157

Qy       1187 tttcaaattgggattcgcagaggcattttggaacatctctggaaaaaattgcaggaccc 1246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 tttcaaattgggattcgcagaggcattttggaacatctctggaaaaaattgcaggaccc 1217

Qy       1247 aagtaatcctgccatcatcaggcaggctgctggaaattatattggaagcttttggcaag 1306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1218 aagtaatcctgccatcatcaggcaggctgctggaaattatattggaagcttttggcaag 1277

Qy       1307 agctaaatttattcctcttattactgtaaaatcatgcctagatcttttggttaactggct 1366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1278 agctaaatttattcctcttattactgtaaaatcatgcctagatcttttggttaactggct 1337

Qy       1367 gcacatatacctaataaccaggattcgggaacaaaggcattctgcgatgttgctctcca 1426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 gcacatatacctaataaccaggattcgggaacaaaggcattctgcgatgttgctctcca 1397

Qy       1427 tggaccattttactcagcctgccaagctgtgttctacacctttgttttagacacaagca 1486
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1398 tggaccattttactcagcctgccaagctgtgttctacncctttgttttagacacaagca 1457

Qy       1487 gcttttgagcggaaacctgaaagaaggtttgcagtatcttcagagtctgaattttgagcg 1546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1458 gcttttgagcggaaacctgaaagaaggtttgcagtatcttcagagtctgaattttgagcg 1517

Qy       1547 gatagtgatgagccagctaaatcccctgaagatttgcctgccctcagtggttaactttt 1606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1518 gatagtgatgagccagctaaatcccctgaagatttgcctgccctcagtggttaactttt 1577
```

```
Qy   1607 tgctgcaatcacaaataagtaccagctcgtcttctgctacaccatcattgaqaqgaacaa 1666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1578 tgctgcaatcacaaataagtaccagctcgtcttctgctacaccatcattgaqaqgaacaa 1637

Qy   1667 tcgccaqatgctgccagtcattaqgaqtaccgctggaqgaqactcagtgcagatctgcac 1726
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1638 tcgccaqatgctgccagtcattaqgaqtaccgctggaqgaqactcagtgcagatctgcac 1697

Qy   1727 aaacccgctggacacctttcttcccctttgatccctgtgtgctgaagaggtcaaaqaaatt 1786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1698 aaacccgctggacacctttcttcccctttgatccctgtgtgctgaagaggtcaaaqaaatt 1757

Qy   1787 cattgatcctatttatcaggtatgggaagacatgaqtgctgaagaqctacaggagttcaa 1846
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1758 cattgatcctatttatcaggtgtgggaagacatgaqtgctgaagaqctacaggagttcaa 1817

Qy   1847 qaaacccatgaaaaaqgacataqtggaagatgaagatgatgacttttctgaaaggcgaagt 1906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1818 qaaacccatgaaaaaggacataqtggaagatgaagatgatgacttttctgaaaggcgaagt 1877

Qy   1907 qcccqaqaatgataccgtgattgggatcacaccaaqctcctttgacacgcatttccgaaq 1966
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1878 qcccqaqaatgataccgtgattgggatcacaccaaqctcctttgacacgcatttccgaaq 1937

Qy   1967 tccttcaaqtaqtgtgggctcccqacccqtgttgtacatgcaaccqagtcccctctqacq 2026
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1938 tccttcaaqtaqtgtgggctcccqacccqtgttgtacatgcaaccqagtcccctctqacq 1997

Qy   2027 qcaqaaatttgtgactqaqatgtgacatttqqqattccccat 2068
          |||||||||||||||||||||||||||||||||||||||||
Db   1998 qcaqaaatttgtgactgagatgtgacatttqqqattccccat 2039
```

RESULT 2
AAC76937
ID AAC76937 standard; cDNA; 1418 BP.
XX
AC AAC76937;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2492 polynucleotide sequence SEQ ID NO:4983.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR P-PSDB; AAB42728.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 4157-4158; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 1418 BP; 422 A; 275 C; 318 G; 403 T; 0 other;

```
  Query Match           44.5%;  Score 920;  DB 21;  Length 1418;
  Best Local Similarity 99.6%;  Pred. No. 0;
  Matches 1120;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    781 cettgaqgcatgaaattctqqaqcttattattgaaaaactactcaaqttqqatqtqaatq 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113 cettgaggcatgaaattctqqaqcttattattgaaaaactactcaaqttqqatqtqaatq 172

Qy    841 catcccqqcaqqqtattqaaqatqctqaaqaaacaqcaactcaaaacttqtqqtqqqacaq 900
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    173 catcccqqcaqqqtattqaaqatqctqaaqaaacaqcaactcaaaacttttqqtqqqacaq 232

Qy    901 attccacqqaaqcattgtttaatatqqatqaaqatqaaqaaactqaacatqaaacaaaqq 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    233 attccacqqaaqcattgtttaatatqqatqaaqatqaaqaaactqaacatqaaacaaaqq 292

Qy    961 ctqqtcctqaacqqctcqaccagatqqtqcatcctqtaqccqaqcqcctqqacatcctqa 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    293 ctqgtcctgaacqqctcgaccagatqgtgcatcctgtagccgagcgcctgqacatcctqa 352

Qy   1021 tqtctttqqttttqtcctacatqaaqqatqtctqctatqtaqatqqtaaqqttqataacq 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    353 tgtctttggttttgtcctacatgaaggatgtctgctatgtagatggtaaggttgataacg 412

Qy   1081 qcaaaacaaaqqatctatatcqcqacctqataaacatctttqacaaactcctqttqccca 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    413 gcaaaacaaaggatctatatcgcgacctgataaacatctttgacaaactcctgttgccca 472

Qy   1141 ccatqcctcctqccatqtacaqttttcatqttttacctctqtaqtttcaaattqqqat 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    473 ccatgcctcctgccatgtacagttttcatgttttacctctgtagtttcaaattgggat 532

Qy   1201 tcqcaqaqqcatttttqqaacatctctqqaaaaaattqcaqqaccccaaqtaatcctqcca 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    533 tcgcagaggcatttttggaacatctctggaaaaaattgcaggaccccaagtaatcctgcca 592

Qy   1261 tcatcaqccaqqctqctqqaaattatattqqaaqcttttqqcaaqaqctaaatttattc 1320
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      593 tcatcagggcaggctgctggaaattatattggaagcttttggcaagagctaaatttattc 652

Qy     1321 ctcttattactgtaaaatcatgcctagatctttggttaactggctgcacatatacctta 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      653 ctcttattactgtaaaatcatgcctagatctttggttaactggctgcacatatacctta 712

Qy     1381 ataaccaggattcgggaacaaaggcattctgcgatgttgctctccatggaccattttact 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      713 ataaccaggattcgggaacaaaggcattctgcgatgttgctctccatggaccattttact 772

Qy     1441 cagcctgccaagctgtgttctacacctttgttttagacacaagcagctttgagcggaa 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      773 cagcctgccaagctgtgttctacacctttgttttagacacaagcagctttgagcggaa 832

Qy     1501 acctgaaagaaggtttgcagtatcttcagagtctgaatttttgagcggatagtgatgagcc 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      833 acctgaaagaaggtttgcagtatcttcagagtctgaatttttgagcggatagtgatgagcc 892

Qy     1561 agctaaatcccctgaagatttgcctgccctcagtggttaacttttttgctgcaatcacaa 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      893 agctaaatcccctgaagatttgcctgccctcagtggttaacttttttgctgcaatcacaa 952

Qy     1621 ataagtaccagctcgtcttctgctacaccatcattgagaggaacaatcgccagatgctgc 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      953 ataagtaccagctcgtcttctgctacaccatcattgagaggaacaatcgccagatgctgc 1012

Qy     1681 cagtcattaggagtaccgctggaggagactcagtgcagatctgcacaaacccgctggaca 1740
            ||||||||||||||||||||||||||||||||||||||| |||||||||||| |||||||
Db     1013 cagtcattaggagtaccgctggaggagactcagtgcagacctgcacaaacccactggaca 1072

Qy     1741 ccttcttccccttgatccctgtgtgctgaagaggtcaaagaaattcattgatcctattt 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1073 ccttcttccccttgatccctgtgtgctgaagaggtcaaagaaattcattgatcctattt 1132

Qy     1801 atcaggtatgggaagacatgaagtgctgaagagctacaggagttcaagaaacccatgaaaa 1860
            ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1133 atcaggtgtgggaagacatgaagtgctgaagagctacaggagttcaagaaacccatgaaaa 1192

Qy     1861 aggacatagtggaagatgaagatgatgactttctgaaaggcgaa 1904
            ||||||||||||||||||||||||||||||||||||:|||||||
Db     1193 aggacatagtggaagatgaagatgatgactttctgaaaggcgaa 1236
```

RESULT 3
AAS44584
ID AAS44584 standard; DNA; 1498 BP.
XX
AC AAS44584;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human full-length polynucleotide sequence #9.
XX
KW Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW nervous system disorder; inflammatory disorder; cell differentiation; ds;
KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;
KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;
KW immunostimulant; analgesic; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200164834-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04926.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
PR 17-JUN-2000; 2000US-0597707.
PR 14-JUL-2000; 2000US-0616807.
PR 19-SEP-2000; 2000US-0664641.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
PI Drmanac R;
XX
DR WPI; 2001-589862/66.
DR P-PSDB; AAU27684.
XX
PT Novel polypeptides and nucleic acids obtained from cDNA libraries
PT prepared from various human tissues, for diagnosis, treatment of
PT cancer, neurological, inflammatory disorders and for use in arrays for
PT detection -
XX
PS Claim 1; SEQ ID No 9; 154pp; English.
XX
CC Sequences AAS44576-AAS44919 represent full length polynucleotides and
CC contig polynucleotides encoding polypeptides of the invention. The DNA
CC and protein sequences are useful for the treatment, diagnosis and
CC prevention of various types of disorder in a mammalian subject such as a
CC human, dog, monkey, mouse, hamster or rat. The disorders include cancers
CC such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such
CC as multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC bowel disease. The sequences exhibit activity relating to angiogenesis,
CC cell proliferation, cell differentiation, stem cell growth factor,
CC activin or inhibin. Therefore, they can be used to manipulate stem cells
CC in culture to give rise to neuroepithelial cells that can be used to
CC augment or replace cells damaged by illness, accidental damage or genetic
CC disorders. The sequences may also be used for regeneration of bone,
CC cartilage, tendons and ligaments and in tissue repair and burn healing.
CC Note: Some sequences for this patent did not form part of the printed
CC specification, but were obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1498 BP; 417 A; 332 C; 325 G; 424 T; 0 other;

```
  Query Match             40.7%;  Score 842;  DB 22;  Length 1498;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  842;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      813 gaaaaactactcaagttggatgtgaatgcatcccggcagggtattgaagatgctgaagaa 872
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       81 gaaaaactactcaagttggatgtgaatgcatcccggcagggtattgaagatgctgaagaa 140

Qy      873 acagcaactcaaacttgtggtgggacagattccacggaaggattgtttaatatggatgaa 932
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      141 acagcaactcaaacttgtggtgggacagattccacggaaggattgtttaatatggatgaa 200

Qy      933 gatgaagaaactgaacatgaaacaaaggctggtcctgaacggctcgaccagatggtgcat 992
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      201 gatgaagaaactgaacatgaaacaaaggctggtcctgaacggctcgaccagatggtgcat 260

Qy      993 cctgtagccgagcgcctggacatcctgatgtctttggtttttgtcctacatgaaggatgtc 1052
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      261 cctgtagccgagcgcctggacatcctgatgtctttggtttttgtcctacatgaaggatgtc 320

Qy     1053 tgctatgtagatggtaaggttgataacggcaaaacaaaggatctatatcgcgacctgata 1112
            |||||||·||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      321 tgctatgtagatggtaaggttgataacggcaaaacaaaggatctatatcgcgacctgata 380

Qy     1113 aacatctttgacaaactcctgttgcccacccatgcctcctgccatgtacagttttcatg 1172
```

```
Db     381 aacatctttgacaaactcctgttgcccacccatgcctcctgccatgtacagttttttcatg 440

Qy    1173 ttttacctctgtagtttcaaattgggattcgcagaggcatttttttgggaacatctctggaaaa 1232
Db     441 ttttacctctgtagtttcaaattgggattcgcagaggcatttttttgggaacatctctggaaaa 500

Qy    1233 aaattgcaggacccaagtaatcctgccatcatcaggcaggctgctgggaaattatattggga 1292
Db     501 aaattgcaggacccaagtaatcctgccatcatcaggcaggctgctgggaaattatattggga 560

Qy    1293 agcttttttggcaagaggctaaatttattcctcttattactgtaaaatcatgcctagatctt 1352
Db     561 agcttttttggcaagaggctaaatttattcctcttattactgtaaaatcatgcctagatctt 620

Qy    1353 ttgttaactggctgcacatatacctaataaccaggattcgggaacaaaggcattctgc 1412
Db     621 ttgttaactggctgcacatatacctaataaccaggattcgggaacaaaggcattctgc 680

Qy    1413 gatgttgctctccatggaccatttttactcagcctgccaagctgtgttctacacctttgtt 1472
Db     681 gatgttgctctccatggaccatttttactcagcctgccaagctgtgttctacacctttgtt 740

Qy    1473 tttagacacaagcagctttgaagcggaaacctgaaagaaggtttgcagtatcttcagagt 1532
Db     741 tttagacacaagcagctttgaagcggaaacctgaaagaaggtttgcagtatcttcagagt 800

Qy    1533 ctgaattttgagcggataqtgatgagccagctaaatcccctgaagatttgcctgcccctca 1592
Db     801 ctgaattttgagcggataqtgatgagccagctaaatcccctgaagatttgcctgcccctca 860

Qy    1593 gtggttaacttttttgctgcaatcacaaataagtaccagctcgtcttctgctacaccatc 1652
Db     861 gtggttaacttttttgctgcaatcacaaataagtaccagctcgtcttctgctacaccatc 920

Qy    1653 at 1654
Db     921 at 922
```

RESULT 4
AAS01562
ID AAS01562 standard; cDNA; 1423 BP.
XX
AC AAS01562;
XX
DT 18-JUL-2001 (first entry)
XX
DE Human secretory molecule cDNA sptm #52.
XX
KW Human; secretory molecule; sptm; SPTM; library screening; gene therapy;
KW cell signalling; cell proliferative disorder; atherosclerosis; cancer;
KW immune system disorder; AIDS; neurological disorder; Alzheimer's disease;
KW nervous system disease; mental retardation; developmental disorder;
KW neuromuscular disorder; microarray; Incyte ID number 4287452dec; ss.
XX
OS Homo sapiens.
XX
PN WO200123558-A2.
XX
PD 05-APR-2001.
XX
PF 19-SEP-2000; 2000WO-US25610.
XX
PR 28-SEP-1999; 99US-0156624.
PR 28-SEP-1999; 99US-0156625.
PR 02-DEC-1999; 99US-0168611.
PR 02-DEC-1999; 99US-0168613.
PR 02-DEC-1999; 99US-0168614.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Hodgson DM, Lincoln SE, Russo FD, Spiro PA, Banville SC;
PI Britcher SR, Dufour GE, Cohen BJ, Rosen BH, Shah P, Chalup MS;
PI Hillman JL, Jones AL, Yu JY, Greenawalt LB, Panzer SR;
PI Roseberry AM, Wright RJ, Chen W, Liu TF, Yap PE, Stockdreher IK;
PI Amshey S, Fong WT;
XX
DR WPI; 2001-258134/26.
XX
PT New secretory polynucleotides (sptm) and the polypeptides they encode,
PT for use in inducing antibodies and screening libraries of compounds -
XX
PS Claim 1; Page 152; 161pp; English.
XX
CC The present sequence for human secretory molecule cDNA sptm #52
CC (Incyte ID number 4287452dec) is 1 of 63 novel sptm cDNA sequences
CC (AAS01511-AAS01573) which encode for the secretory polypeptides SPTM. The
CC sptm polynucleotides are useful for screening a compound for
CC effectiveness in altering expression of a target polynucleotide, where
CC the target polynucleotide comprises sptm. Sptm is also useful in a method
CC for assessing the toxicity of a test compound. Sptm and its fragments or
CC complementary sequences, may be used to identify the presence of and/or
CC determine the degree of similarity between two nucleic acid sequences.
CC Sptm can also be used for a variety of diagnostic and therapeutic
CC purposes, e.g. diagnosing a particular condition, disease or disorder
CC associated with cell signalling, such conditions include cell
CC proliferative disorders such as atherosclerosis, and cancers including
CC leukaemia, an immune system disorder e.g. acquired immunodeficiency
CC syndrome (AIDS), a neurological disorder such as epilepsy or Alzheimer's
CC disease, nutritional and metabolic disease of the nervous system, mental
CC retardation and other developmental disorders, and muscular dystrophy and
CC other neuromuscular disorders. Sptm can also be used to design probes
CC useful in diagnostic assays, which may be used to monitor the progress of
CC conditions or disorders associated with abnormal levels of expression of
CC sptm. In addition sptm encoding SPTM may be used for somatic or germline
CC gene therapy, for inducing antibodies, or in microarrays.
XX
SQ Sequence 1423 BP; 417 A; 247 C; 320 G; 439 T; 0 other;

Query Match 33.4%; Score 691; DB 22; Length 1423;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 691; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     189 ttttttcaattctcccccaaqaaaaactqttcqqtttqqtqqaactqtqacaqaaqtcttq 248
Db     224 ttttttcaattctcccccaaqaaaaactqttcqqtttqqtqqaactqtqacaqaaqtcttq 283

Qy     249 ctqaaqtacaaaaaqqqtqaaacaaatqactttqaqttqttqaaqaaccaqctqttaqat 308
Db     284 ctqaaqtacaaaaaqqqtqaaacaaatqactttqaqttqttqaaqaaccaqctqttaqat 343

Qy     309 ccaqacataaaqqatqaccaqatcatcaactqqctqctaqaattccqttcttctatcatq 368
Db     344 ccaqacataaaqqatqaccaqatcatcaactqqctqctaqaattccqttcttctatcatq 403

Qy     369 tacttqacaaaaqactttqaqcaacttatcaqtattatattaaqattqccttqqttqaat 428
Db     404 tacttqacaaaaqactttqaqcaacttatcaqtattatattaaqattqccttqqttqaat 463

Qy     429 agaagtcaaacagtagtgqaagaqtatttggcttttcttggtaatcttgtatcaqcacaq 488
Db     464 agaagtcaaacagtagtgqaagaqtatttggcttttcttggtaatcttgtatcaqcacaq 523

Qy     489 actgttttcctcaqaccgtgtctcaqcatgattgcttcccatttgtgcctccccqaqtq 548
Db     524 actgttttcctcaqaccgtgtctcaqcatgattgcttcccatttgtgcctccccqaqtq 583

Qy     549 atcattaaqqaaqqcqatqtaqatqtttcaqattctqatqatqaaqatqataatcttcct 608
Db     584 atcattaaqqaaqqcqatqtaqatqtttcaqattctqatqatqaaqatqataatcttcct 643

Qy     609 qcaaattttqacacatqtcacaqaqccttqcaaataatagcaaqatatqtaccatcaaca 668
Db     644 qcaaattttqacacatqtcacaqaqccttqcaaataatagcaaqatatqtaccatcaaca 703
```

```
Qy     669 ccqtqqtttctcatqccaatactqqtqqaaaaatttccatttqttcqaaaatcaqaqaqa 728
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     704 ccqtqqtttctcatqccaatactqqtqqaaaaatttccatttqttcqaaaatcaqaqaqa 763

Qy     729 acactqqaatqttacqttcataacttactaaqqattaqtqtatatttttccaaccttqaqq 788
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     764 acactqqaatqttacqttcataacttactaaqqattaqtqtatatttttccaaccttqaqq 823

Qy     789 catqaaattctqqaqcttattattqaaaaactactcaaqttqqatqtqaatqcatcccqq 848
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     824 catqaaattctqqaqcttattattqaaaaactactcaaqttqqatqtqaatqcatcccqq 883

Qy     849 caqqtattqaaqatqctqaaqaaacaqcaa 879
           |||||||||||||||||||||||||||||
Db     884 caqqtattqaaqatqctqaaqaaacaqcaa 914


RESULT   5
AAV90248
ID   AAV90248 standard; cDNA; 341 BP.
XX
AC   AAV90248;
XX
DT   15-FEB-1999  (first entry)
XX
DE   EST clone DI449.
XX
KW   Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW   tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW   receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW   gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9845446-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06955.
XX
PR   10-APR-1997;   97US-0838821.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070077/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries.
XX
PS   Claim 1; Page 478; 618pp; English.
XX
CC   The present sequence represents a human expressed sequence tag (EST).
CC   The polynucleotide, which is a secreted EST, and the encoded protein
CC   are predicted to have useful biological activities which would make
CC   them suitable for treating, preventing or ameliorating medical
CC   conditions in humans and animals, although no supporting data is
CC   given. Suggested activities include nutritional activity, immune
CC   stimulating or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The polynucleotide may also be useful for gene therapy.
XX
SQ   Sequence 341 BP; 86 A; 83 C; 80 G; 92 T; 0 other;


  Query Match             14.1%;  Score 292;  DB 20;  Length 341;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-140;
  Matches  292;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1485 caqctttqaqcqgaaacctqaaaqaaqgttqcaqtatcttcaqaqtctqaattttqaq 1544
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22 caqctttgaqcggaaacctqaaaqaaqgtttqcaqtatcttcaqaqtctqaattttqaq 81

Qy    1545 cqqataqtqatqaqccaqctaaatcccctqaaqatttqcctqccctcaqtqqttaacttt 1604
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82 cqqataqtqatqaqccaqctaaatcccctqaaqatttqcctqccctcaqtqqttaacttt 141

Qy    1605 tttqctqcaatcacaaataaqtaccaqctcqtcttctqctacaccatcattqaqaqqaac 1664
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142 tttqctqcaatcacaaataaqtaccaqctcqtcttctqctacaccatcattqaqaqqaac 201

Qy    1665 aatcqccaqatqctqccaqtcattaqqaqtaccqctqqaqqaqactcaqtqccaqatctqc 1724
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     202 aatcqccaqatqctqccaqtcattaqqaqtaccqctqqaqqaqactcaqtqccaqatctqc 261

Qy    1725 acaaacccqctqqacacttcttcccctttqatccctatqtqctqaaqaqqt 1776
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     262 acaaacccqctqqacacttcttcccctttqatccctatqtqctqaaqaqqt 313


RESULT   6
AAS87641
ID   AAS87641 standard; cDNA; 3169 BP.
XX
AC   AAS87641;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #23445.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/74.
DR   P-PSDB; AB023454.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 23445; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
```

```
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3169 BP; 812 A; 841 C; 779 G; 737 T; 0 other;


  Query Match             13.0%;  Score 269;  DB 23;  Length 3169;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-119;
  Matches  269;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1245 ccaaqtaatcctqccatcatcaqqcaqqctqctqqaaattatattqqaaqctttttqqca 1304
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    878 ccaaqtaatcctqccatcatcaqqcaqqctqctqqaaattatattqqaaqctttttqqca 937

Qy   1305 aqaqctaaatttattcctcttattactqtaaaatcatqcctaqatcttttqqttaactqq 1364
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    938 aqaqctaaatttattcctcttattactqtaaaatcatqcctaqatcttttqqttaactqq 997

Qy   1365 ctqcacatatacccttaataaccaqqattcqqqaacaaaqqcattctqcqatqttqctctc 1424
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    998 ctqcacatatacccttaataaccaqqattcqqqaacaaaqqcattctqcqatqttqctctc 1057

Qy   1425 catqqaccattttactcaqcctqccaaqctqtqttctacacctttqtttttaqacacaaq 1484
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1058 catqqaccattttactcaqcctqccaaqctqtqttctacacctttqtttttaqacacaaq 1117

Qy   1485 caqcttttqaqcqqaaacctqaaaqaaqq 1513
           |||||||||||||||||||||||||||||
Db   1118 caqcttttqaqcqqaaacctqaaaqaaqq 1146


RESULT   7
ABA83038
ID   ABA83038 standard; DNA; 1461 BP.
XX
AC   ABA83038;
XX
DT   05-FEB-2002  (first entry)
XX
DE   Human transcription factor TRFX-65 coding sequence.
XX
KW   Human; transcription factor; TRFX; cell proliferative disease;
KW   autoimmune disease; inflammation; neurological disease;
KW   developmental disorder; cancer; AIDS; infection; cytostatic; anti-HIV;
KW   neuroprotective; antiinflammatory; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200172777-A2.
XX
PD   04-OCT-2001.
XX
PF   13-MAR-2001; 2001WO-US08117.
XX
PR   13-MAR-2000; 2000US-0188986.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Hillman JL,  Baughn MR,  Yue H,  Lal P,  Lu DAM,  Patterson C;
PI   Azimzai Y,  Bandman O,  Tang YT,  Mathur P,  Shah P,  Au-Young J;
PI   Reddy R;
XX
DR   WPI; 2001-570896/64.
DR   P-PSDB; ABB50214.
XX
PT   Novel transcription factor polypeptides, used to treat diseases
PT   associated with altered activity and expression of TRFX, and to screen
PT   for agents capable of modulating its activity -
XX
PS   Claim 11; Page 299; 327pp; English.
XX
CC   The present sequence is the coding sequence for a human transcription
CC   factor. The transcription factor and its coding sequence are useful in
CC   the diagnosis, treatment and prevention of diseases associated with
CC   altered expression of the transcription factor e.g. cell proliferative,
CC   autoimmune/inflammatory, neurological and developmental disorders. A
CC   number of specific disorders/diseases are given in the specification,
CC   including: arteriosclerosis, cirrhosis, hepatitis, cancers, AIDS,
CC   allergies, anaemia, asthma, autoimmune thyroiditis, bronchitis, atopic
CC   dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,
CC   Grave's disease, multiple sclerosis, osteoarthritis, pancreatitis,
CC   psoriasis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative
CC   colitis, uveitis, Alzheimer's disease, Huntington's disease, Parkinson's
CC   disease, stroke, and viral, bacterial, fungal and protozoal infections.
XX
SQ   Sequence 1461 BP; 413 A; 335 C; 298 G; 415 T; 0 other;


  Query Match             11.2%;  Score 232;  DB 22;  Length 1461;
  Best Local Similarity   99.6%;  Pred. No. 2.8e-101;
  Matches  282;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1339 catqcctaqatcttttqqttaactqqctqcacatatacccttaataaccaqqattcqqqaaa 1398
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    681 catqcctaqatcttttqqttaactqqctqcacatatacccttaataaccaqqattcqqqaaa 740

Qy   1399 caaaqqcattctqcqatqttqctctccatqqaccattttactcaqcctqccaaqctqtqt 1458
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    741 caaaqqcattctqcqatqttqctctccatqqaccattttactcaqcctqccaaqctqtqt 800

Qy   1459 tctacacctttqtttttaqacacaaqcaqcttttqaqcqqaaacctqaaaqaaqqtttqc 1518
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    801 tctacacctttqtttttaqacacaaqcaqcttttqaqcqqaaacctqaaaqaaqqtttqc 860

Qy   1519 aqtatcttcaqaqtctqaatttqaaqcqqatagtqatqaqccaqctaaatccccqaaqa 1578
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    861 aqtatcctcaqaqtctqaatttqaaqcqqatagtqatqaqccaqctaaatccccqaaqa 920

Qy   1579 tttqcctqcccqcaqtqqttaacttttttqctqcaatcacaaa 1621
           |||||||||||||||||||||||||||||||||||||||||||
Db    921 tttqcctqcccqcaqtqqttaacttttttqctqcaatcacaaa 963


RESULT   8
AAC77798
ID   AAC77798 standard; cDNA; 2103 BP.
XX
AC   AAC77798;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human cancer associated gene sequence SEQ ID NO:192.
XX
KW   Human; cancer associated gene; cancer antigen; detection; cancer;
KW   diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW   antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW   antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW   dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW   vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW   immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW   allergic reaction; graft versus host disease; organ rejection;
KW   haemostatic; thrombolytic; cardiovascular disorder; infection;
KW   neurological disease; drug screening; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055350-A1.
XX
```

```
PD   21 SEP 2000.
XX
PF   08 MAR 2000; 2000WO US05882.
XX
PR   12 MAR 1999;  99US 0124270.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587533/55.
DR   P-PSDB; AAB44589.
XX
PT   Novel isolated nucleic acids comprising sequences encoding peptides
PT   useful for treating or diagnosing e.g. cancer
XX
PS   Claim 1; Page 767 768; 2452pp; English.
XX
CC   AAC77607 to AAC78448 encode the human cancer associated proteins given
CC   in AAB43998 to AAB44239. The proteins can have activities based on the
CC   tissues and cells the genes are expressed in. Example of activities
CC   include: cytostatic; proliferative; vulnerary; immunomodulator;
CC   antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC   antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC   dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC   nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC   polynucleotides and polypeptides can be used for preventing, treating or
CC   ameliorating medical conditions and diagnosing pathological conditions.
CC   Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC   the present invention may be used to treat immune disorders by activating
CC   or inhibiting the proliferation, differentiation or mobilisation of
CC   immune cells, to treat disorders of haematopoietic cells, autoimmune
CC   disorders, allergic reactions, graft versus host disease and organ
CC   rejection, modulate haemostatic or thrombolytic activity, modulate
CC   inflammation, cancers, cardiovascular disorders, neurological disease and
CC   bacterial or viral infections. The peptides, nucleotides, antibodies,
CC   agonists and antagonists may be also be used in drug screens. AAC78449 to
CC   AAC78457 and AAB44240 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 2103 BP; 596 A; 345 C; 486 G; 674 T; 2 other;


  Query Match             11.1%;  Score 230;  DB 21;  Length 2103;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-100;
  Matches  230;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1869 gaqttcaaqaaaaccatqaaaaaqqacataqtqqaaqatqaaqatqatqacttttctqaaa 1928
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     179 gaqttcaaqaaaaccatqaaaaaqqacataqtqqaaqatqaaqatqatqacttttctqaaa 238

Qy    1929 qqcqaaqtqccccaqaatqataccqtqattqqqatcacaccaaqctccttqacacqcat 1988
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     239 qqcqaaqtqccccaqaatqataccqtqattqqqatcacaccaaqctccttqacacqcat 298

Qy    1989 ttccqaaqtccttcaaqtaqtqtqqqctccccaccqtqttqtacatqccaaccccaqtccc 2048
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     299 ttccqaaqtccttcaaqtaqtqtqqqctccccaccqtqttqtacatqccaaccccaqtccc 358

Qy    2049 ctctqacqqccaqaaatttqtqactqaqatqtqacatttqqqattccccat 2098
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     359 ctctqacqqccaqaaatttqtqactqaqatqtqacatttqqqattccccat 408


RESULT   9
AAS73660
ID   AAS73660 standard; cDNA; 2991 BP.
XX
AC   AAS73660;
XX
DT   13 FEB 2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #9464.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-689362/78.
DR   P-PSDB; ABG09473.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 9464; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2991 BP; 655 A; 853 C; 882 G; 601 T; 0 other;


  Query Match             11.1%;  Score 229;  DB 23;  Length 2991;
  Best Local Similarity   99.6%;  Pred. No. 7.6e-100;
  Matches  279;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     321 qatqaccaqatcatcaactqqctqctaqaattccqttcttctatcatqtacttqacaaaa 380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     610 aatqaccaqatcatcaactqqctqctaqaattccqttcttctatcatqtacttqacaaaa 669

Qy     381 qactttqaqcaacttatcaqtattatattaaqattqccttqqttqaataqaaaqtcaaaca 440
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     670 qactttqaqcaacttatcaqtattatattqaqattqccttqqttqaataqaaaqtcaaaca 729

Qy     441 atagtqqaaaqaqtatttqqctttcttqqtaatcttqtatcaqcacaqactqttttccctc 500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     730 atagtqqaaaqaqtatttqqctttcttqqtaatcttqtatcaqcacaqactqttttccctc 789

Qy     501 aqaccqtqtctcaqcatqattqcttcccatttqtqcctccccqaqtqatcattaaqqaa 560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     790 aqaccqtqtctcaqcatqattqcttcccatttqtqcctccccqaqtqatcattaaqqaa 849
```

```
Qy     561 ggcgatgtagatgtttcagattctgatgatgaagatgata 600
           ||||||||||||||||||||||||||||||||||||||||||
Db     850 ggcgatgtagatgtttcagattctgatgatgaagatgata 889


RESULT  10
AAS91992
ID   AAS91992 standard; cDNA; 2991 BP.
XX
AC   AAS91992;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #27796.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG27805.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 1; SEQ ID No 27796; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2991 BP; 655 A; 853 C; 882 G; 601 T; 0 other;

  Query Match             11.1%;  Score 229;  DB 23;  Length 2991;
  Best Local Similarity   99.6%;  Pred. No. 7.6e-100;
  Matches  279;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     321 gatgaccagatcatcaactgactgctagaattccgttcttctatcatgtacttgacaaaa 380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     610 gatgaccagatcatcaactgactgctagaattccgttcttctatcatgtacttgacaaaa 669

Qy     381 gactttgaagcaacttatcagtattatattaagattgccttgattgaataaaatcaaaca 440
           ||||||||||||||||||||||||||||| ||||||||||||||||||||,,,,,!!!!!
Db     670 gactttgagcaacttatcagtattatattgaagattgccttgattgaataaaatcaaaca 729

Qy     441 gtaatgaaaagaatatttgactttttcttagtaatcttgtatcaagcacaaactgttttcctc 500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     730 gtaatgaaaagaatatttgactttttcttagtaatcttgtatcaagcacaaactgttttcctc 789

Qy     501 agaccgtgtctcagcatgattgcttcccatttttgtgcctccccgaagtgatcattaaagaaa 560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     790 agaccgtgtctcagcatgattgcttcccatttttgtgcctccccgaagtgatcattaaagaaa 849

Qy     561 ggcgatgtagatgtttcagattctgatgatgaagatgata 600
           ||||||||||||:||||||||||||||||||||||||||
Db     850 ggcgatgtagatgtttcagattctgatgatgaagatgata 889


RESULT  11
AAS92253
ID   AAS92253 standard; cDNA; 776 BP.
XX
AC   AAS92253;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #28057.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG28066.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 1; SEQ ID No 28057; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
```

```
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 776 BP; 206 A; 157 C; 198 G; 215 T; 0 other;


  Query Match             10.8%;  Score 223;  DB 23;  Length 776;
  Best Local Similarity   100.0%;  Pred. No. 6e-97;
  Matches  223;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     419 ttqqttqaataqaaqtcaaacaqtaqtqqaaaqaqtatttqqctttttcttqqtaatcttqt 478
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     553 ttqqttqaataqaaqtcaaacaqtaqtqqaaaqaqtatttqqctttttcttqqtaatcttqt 612

Qy     479 atcaqcacaqactqtttttcctcaqaccqtqtctcaqcatqattqcttcccatttqtqcc 538
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     613 atcaqcacaqactqtttttcctcaqaccqtqtctcaqcatqattqcttcccatttqtqcc 672

Qy     539 tcccqaqtqatcattaaqqaaqqcqatqtaqatqtttcaqattctqatqatqaaqatqa 598
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     673 tcccqaqtqatcattaaqqaaqqcqatqtaqatqtttcaqattctqatqatqaaqatqa 732

Qy     599 taatcttcctqcaaatttqacacatqtcacaqaqccttqcaa 641
           |||||||||||||||||||||||||||||||||||||||||||
Db     733 taatcttcctqcaaatttqacacatqtcacaqaqccttqcaa 775


RESULT  12
AAS92254
ID   AAS92254 standard; cDNA; 3175 BP.
XX
AC   AAS92254;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #28058.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001
XX
PF   30-MAR-2001; 2001WO-US08641.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   24-AUG-2000; 2000US-0649167.
XX
PA   (HYSE ) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-649362/73.
DR   P-PSDB; AAG28067.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 1; SEQ ID No 28058; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3175 BP; 811 A; 849 C; 779 G; 736 T; 0 other;


  Query Match             10.4%;  Score 215;  DB 23;  Length 3175;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-93;
  Matches  215;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     665 qacacctqqttttctcatqccaatactqqtqqaaaaattccatttqttcqaaaatcaqa 724
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     490 qacacctqqttttctcatqccaatactqqtqqaaaaattccatttqttcqaaaatcaqa 549

Qy     725 qaqaacactqqaatqttacqttcataacttactaaqqattaqtqtatatttccaacctt 784
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     550 qaqaacactqqaatqttacqttcataacttactaaqqattaqtqtatatttccaacctt 609

Qy     785 qaqqcatqaaattctqqaqcttattatttqaaaaactactcaaqttqqatqtqaatqcatc 844
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     610 qaqqcatqaaattctqqaqcttattatttqaaaaactactcaaqttqqatqtqaatqcatc 669

Qy     845 ccqqcaqqqtattqaaqatqctqaaqaaacaqccaa 879
           |||||||||||||||||||||||||||||||||||
Db     670 ccqqcaqqqtattqaaqatqctqaaqaaacaqccaa 704


RESULT  13
AAK68677
ID   AAK68677 standard; DNA; 34269 BP.
XX
AC   AAK68677;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23489.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198124.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
```

PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -
XX
PS Disclosure; SEQ ID NO 23489; 3071pp - Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased

```
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic related diseases, especially
CC   cancers and cancer metastases of haematopoietic derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 34269 BP; 9052 A; 6778 C; 7638 G; 10801 T; 0 other;


  Query Match             10.1%;  Score 208;  DB 22;  Length 34269;
  Best Local Similarity   100.0%;  Pred. No. 9.3e-90;
  Matches  208;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1861 aqqacataqtqqaaqatqaaqatqatqacttttctqaaaqqcqaaqtqccccaqaatqata 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   32384 aqqacataqtqqaaqatqaaqatqatqacttttctqaaaqqcqaaqtqccccaqaatqata 32443

Qy    1921 ccqtqattqqqatcacaccaaqctcctttqacacqcattcccqaaqtcctcaaqtaqtq 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   32444 ccqtqattqqqatcacaccaaqctcctttqacacqcattcccqaaqtcctcaaqtaqtq 32503

Qy    1981 tqqqctccccacccqtqttqtacatqcaacccaqtcccctctqacqqccaqaaattttqtqa 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   32504 tqqqctccccacccqtqttqtacatqcaacccaqtcccctctqacqqccaqaaattttqtqa 32563

Qy    2041 ctqaqatqtqacatttqqqattcccat 2068
           |||||||||||||||||||||||||||
Db   32564 ctqaqatqtqacatttqqqattcccat 32591


RESULT  14
AAK85168
ID   AAK85168 standard; DNA; 34269 BP.
XX
AC   AAK85168;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39980.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229514.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   27-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236470.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
```

PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -
XX
PS Disclosure; SEQ ID NO 39980; 307lpp - Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patients own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/haematopoietic-related diseases, especially cancers and cancer metastases of haematopoietic-derived cells. AAK64703 to AAK87694 represent human immune/haematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54950 and AAM82169 represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 34269 BP; 9052 A; 6778 C; 7638 G; 10801 T; 0 other;

Query Match 10.1%; Score 208; DB 22; Length 34269;
Best Local Similarity 100.0%; Pred. No. 9.3e-90;
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1861 aggacataqtqqaaqatqaaqatqatqactttctqaaaqqcqaaqtqccccaqaatqata 1920
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 32384 aggacataqtqqaaqatqaaqatqatqactttctqaaaqqcqaaqtqccccaqaatqata 32443

Qy 1921 ccqtqattqqqatcacaccaaqctcctttqacacqcatttccqaaatcctttcaaqtaqtq 1980
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 32444 ccqtqattqqqatcacaccaaqctcctttqacacqcatttccqaaatcctttcaaqtaqtq 32503

Qy 1981 tqqqctccccacccqtqttqtacatqcaaccqaqtcccctctqacqqcaqaaaatttqtqa 2040
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 32504 tqqqctccccacccqtqttqtacatqcaaccqaqtcccctctqacqqcaqaaaatttqtqa 32563

Qy 2041 ctqaqatqtqacatttqqqattcccat 2068
|||||||||||||||||||||||||||||
Db 32564 ctqaqatqtqacatttqqqattcccat 32591

RESULT 15
AAS92255
ID AAS92255 standard; cDNA; 2410 BP.
XX
AC AAS92255;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #28059.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic; KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; AAG28068.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations PT responsible for genetic disorders or other traits and to assess PT biodiversity -
XX
PS Claim 1; SEQ ID No 28059; 104pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and CC polypeptide (II) sequences. (I) is useful as hybridisation probes, CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome CC and gene mapping, and in recombinant production of (II). The CC polynucleotides are also used in diagnostics as expressed sequence tags CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2410 BP; 672 A; 437 C; 547 G; 754 T; 0 other;

Query Match 9.6%; Score 198; DB 23; Length 2410;
Best Local Similarity 100.0%; Prod. No. 6.2e-85;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1347 qatcttttqqttaactqqctqcacatataccttaataaccaqqattcqqqaacaaaqqca 1406
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 qatcttttqqttaactqqctqcacatataccttaataaccaqqattcqqqaacaaaqqca 60

Qy    1407 ttctqcqatqttqctctccatqqaccatttactcaqcctqccaaqctqtqttctacacc 1466
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ttctqcqatqttqctctccatqqaccatttactcaqcctqccaaqctqtqttctacacc 120

Qy    1467 tttqttttqqacacaaqcaqctttqaqcqqaaacctqaaaqaaqtttqcaqtatctt 1526
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 tttqttttqqacacaaqcaqctttqaqcqqaaacctqaaaqaaqtttqcaqtatctt 180

Qy    1527 caqaqtctqaattttqaq 1544
           ||||||||||||||||||
Db     181 caqaqtctqaattttqaq 198
```

Search completed: July 30, 2002, 05:18:48
Job time: 47.9 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 03:10:00 ; Search time 2610.22 Seconds
(without alignments)
16579.475 Million cell updates/sec

Title: US-09-932-678-1
Perfect score: 2068
Sequence: 1 acaqaqqctqtqqctqqaaq.....tqacatttyqqattccccat 2068

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: qb_ba:*
2: qb_htq:*
3: qb_in:*
4: qb_om:*
5: qb_ov:*
6: qb_pat:*
7: qb_ph:*
8: qb_pl:*
9: qb_pr:*
10: qb_ro:*
11: qb_sts:*
12: qb_sy:*
13: qb_un:*
14: qb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htq_hum:*
31: em_htq_inv:*
32: em_htq_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2068	100.0	2068	9	AF227156	AF227156 Homo sapi
2	1920	92.8	3756	9	HSA272050	AJ272050 Homo sapi
3	1869	90.4	2040	6	AX036047	AX036047 Sequence
4	308	14.9	2493	9	AK055742	AK055742 Homo sapi
5	283	13.7	1177	9	BC009198	BC009198 Homo sapi
6	232	11.2	1461	6	AX274907	AX274907 Sequence
7	232	11.2	2627	9	BC006441	BC006441 Homo sapi
8	208	10.1	129473	2	AC040158	AC040158 Homo sapi
c 9	208	10.1	143900	2	AC092137	AC092137 Homo sapi
10	187	9.0	175691	2	AC007615	AC007615 Homo sapi
11	187	9.0	191496	2	AC009130	AC009130 Homo sapi
12	187	9.0	208008	2	AC009093	AC009093 Homo sapi
c 13	176	8.5	173166	9	AC092375	AC092375 Homo sapi
c 14	176	8.5	179150	2	AC017077	AC017077 Homo sapi
15	176	8.5	202004	9	HUAF001549	AF001549 Human Chr
16	158	7.6	143900	2	AC092137	AC092137 Homo sapi
17	125	6.0	170611	9	AC106788	AC106788 Homo sapi
18	123	5.9	187643	2	AC092562	AC092562 Papio ham
c 19	119	5.8	191496	2	AC009130	AC009130 Homo sapi
c 20	115	5.6	824	9	HSA343405	AJ343405 Homo sapi
c 21	109	5.3	247331	2	AC097268	AC097268 Pan troql
c 22	97	4.7	293	9	HS151B3F	Z56854 H.sapiens C
23	92	4.4	179150	2	AC017077	AC017077 Homo sapi
c 24	79	3.8	196413	2	AC097269	AC097269 Pan troql
25	69	3.3	78347	2	AC022642	AC022642 Homo sapi
c 26	54	2.6	78347	2	AC022642	AC022642 Homo sapi
c 27	47	2.3	108040	2	AC068150	AC068150 Homo sapi
c 28	47	2.3	171940	2	AC092119	AC092119 Homo sapi
29	47	2.3	173166	9	AC092375	AC092375 Homo sapi
c 30	47	2.3	316296	2	AC092285	AC092285 Homo sapi
c 31	42	2.0	821	9	HSA343406	AJ343406 Homo sapi
c 32	36	1.7	405	11	G27379	G27379 human STS S
33	32	1.5	286	9	HS151B3R	Z56855 H.sapiens C
34	30	1.5	135873	2	AC012547	AC012547 Homo sapi
35	30	1.5	192815	9	AC013444	AC013444 Homo sapi
c 36	26	1.3	461	11	G38441	G38441 SHGC 58117
c 37	26	1.3	181088	2	AC097733	AC097733 Rattus no
38	24	1.2	53814	2	AC107769	AC107769 Mus muscu
c 39	24	1.2	53814	2	AC107769	AC107769 Mus muscu
40	22	1.1	737	9	HSA337682	AJ337682 Homo sapi
41	22	1.1	38147	8	SPAC6B12	Z98531 S.pombe chr
c 42	22	1.1	76416	9	AC005370	AC005370 Homo sapi
43	22	1.1	108793	9	AL137839	AL137839 Human DNA
44	22	1.1	146437	9	AC010489	AC010489 Homo sapi
c 45	22	1.1	155022	2	AC108176	AC108176 Bos tauru

ALIGNMENTS

RESULT 1
AF227156
LOCUS AF227156 2068 bp mRNA linear PRI 29-APR-2000
DEFINITION Homo sapiens RRN3 mRNA, complete cds.
ACCESSION AF227156
VERSION AF227156.1 GI:7670099
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2068)
AUTHORS Moorefield,B., Greene,E.A. and Reeder,R.H.
TITLE RNA polymerase I transcription factor Rrn3 is functionally conserved between yeast and human
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (9), 4724-4729 (2000)
MEDLINE 20243763
REFERENCE 2 (bases 1 to 2068)
AUTHORS Moorefield,B., Greene,E.A. and Reeder,R.H.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2000) Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N., Seattle, WA 98109, USA
FEATURES Location/Qualifiers

```
     source          1..2068
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "16"
                     /map "16p12"
     CDS             69..2024
                     /note "similar to yeast RNA polymerase I transcription
                     factor Rrn3"
                     /codon_start 1
                     /product "RRN3"
                     /protein_id "AAF66160.1"
                     /db_xref "GI:7670100"
                     /translation "MAAPLLHTRLPGDAAASSSAVKKLGASRTGISNMRALENDFFNS
                     PPRKTVRFGGTVTEVLLKYKKGETNDFELLKNQLLDPDIKDDQIINWLLEFRSSIMYL
                     TKDFEQLISIILRLPWLNRSQTVVEEYLAFLGNLVSAQTVFLRPCLSMIASHFVPPRV
                     IIKEGDVDVSDSDDEDDNLPANFDTCHRALQIIARYVPSTPWFLMPILVEKFPFVRKS
                     ERTLECYVHNLLRISVYFPTLRHEILELIIEKLLKLDVNASRQGIEDAEETATQTCGG
                     TDSTEGLFNMDEDEETEHETKAGPERLDQMVHPVAERLDILMSLVLSYMKDVCYVDGK
                     VDNGKTKDLYRDLINIFDKLLLPTHASCHVQFFMFYLCSFKLGFAEAFLEHLWKKLQD
                     PSNPAIIRQAAGNYIGSFLARAKFIPLITVKSCLDLLVNWLHIYLNNQDSGTKAFCDV
                     ALHGPFYSACQAVFYTFVFRHKQLLSGNLKEGLQYLQSLNFERIVMSQLNPLKICLPS
                     VVNFFAAITNKYQLVFCYTIIERNNRQMLPVIRSTAGGDSVQICTNPLDTFFPFDPCV
                     LKRSKKFIDPIYQVWEDMSAEELQEFKKPMKKDIVEDEDDDFLKGEVPQNDTVIGITP
                     SSFDTHFRSPSSSVGSPPVLYMQPSPL"
BASE COUNT      573 a    437 c    485 g    573 t
ORIGIN

  Query Match            100.0%;  Score 2068;  DB 9;  Length 2068;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2068;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 acagaggctgtggctggaaggagctgggcatccggcctgaggcgcagcggtcgcgttagt 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACAGAGGCTGTGGCTGGAAGGAGCTGGGCATCCGGCCTGAGGCGCAGCGGTCGCGTTAGT 60

Qy         61 tcggcccaatggcggcaccgctgcttcacacgcgtttgccgggagatgcggccgcttcgt 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCGGCCCAATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGATGCGGCCGCTTCGT 120

Qy        121 cctctgcagttaagaagctgggcgcgtcgaggactgggatttcaaatatgcgtgcattag 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATATGCGTGCATTAG 180

Qy        181 agaatgactttttcaattctcccccaagaaaaactgttcggtttggtggaactgtgacag 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTTCGGTTTGGTGGAACTGTGACAG 240

Qy        241 aagtcttgctgaagtacaaaaagggtgaaacaaatgactttgagttgttgaagaaccagc 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGC 300

Qy        301 tgttagatccagacataaaggatgaccagatcatcaactggctgctagaattccgttctt 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGTTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTT 360

Qy        361 ctatcatgtacttgacaaaagactttgagcaacttatcagtattatattaagattgcctt 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCCTT 420

Qy        421 ggttgaatagaagtcaaacagtagtggaagagtatttggcttttcttggtaatcttgtat 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTGGTAATCTTGTAT 480

Qy        481 cagcacagactgttttcctcagaccgtgtctcagcatgattgcttcccattttgtgcctc 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGTGCCTC 540

Qy        541 cccgagtgatcattaaggaaggcgatgtagatgtttcagattctgatgatgaagatgata 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGATGATGAAGATGATA 600

Qy        601 atcttcctgcaaattttgacacatgtcacagagccttgcaaataatagcaagatatgtac 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATCTTCCTGCAAATTTTGACACATGTCACAGAGCCTTGCAAATAATAGCAAGATATGTAC 660

Qy        661 catcgacaccgtggtttctcatgccaatactggtggaaaaatttccatttgttcgaaaat 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTCGAAAAT 720

Qy        721 cagagagaacactggaatgttacgttcataacttactaaggattagtgtatattttccaa 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CAGAGAGAACACTGGAATGTTACGTTCATAACTTACTAAGGATTAGTGTATATTTTCCAA 780

Qy        781 ccttgaggcatgaaattctggagcttattattgaaaaactactcaagttggatgtgaatg 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAAACTACTCAAGTTGGATGTGAATG 840

Qy        841 catcccggcagggtattgaagatgctgaagaaacagcaactcaaacttgtggtgggacag 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CATCCCGGCAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGGTGGGACAG 900

Qy        901 attccacggaaggattgtttaatatggatgaagatgaagaaactgaacatgaaacaaagg 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATTCCACGGAAGGATTGTTTAATATGGATGAAGATGAAGAAACTGAACATGAAACAAAGG 960

Qy        961 ctggtcctgaacggctcgaccagatggtgcatcctgtagccgagcgcctggacatcctga 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGGTCCTGAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCGCCTGGACATCCTGA 1020

Qy       1021 tgtctttggttttgtcctacatgaaggatgtctgctatgtagatggtaaggttgataacg 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TGTCTTTGGTTTTGTCCTACATGAAGGATGTCTGCTATGTAGATGGTAAGGTTGATAACG 1080

Qy       1081 gcaaaacaaaggatctatatcgcgacctgataaacatctttgacaaactcctgttgccca 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GCAAAACAAAGGATCTATATCGCGACCTGATAAACATCTTTGACAAACTCCTGTTGCCCA 1140

Qy       1141 cccatgcctcctgccatgtacagtttttcatgttttacctctgtagtttcaaattgggat 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCATGCCTCCTGCCATGTACAGTTTTTCATGTTTTACCTCTGTAGTTTCAAATTGGGAT 1200

Qy       1201 tcgcagaggcatttttggaacatctctggaaaaaattgcaggacccaagtaatcctgcca 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGCAGAGGCATTTTTGGAACATCTCTGGAAAAAATTGCAGGACCCAAGTAATCCTGCCA 1260

Qy       1261 tcatcaggcaggctgctggaaattatattggaagctttttggcaagagctaaatttattc 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAGAGCTAAATTTATTC 1320

Qy       1321 ctcttattactgtaaaatcatgcctagatcttttggttaactggctgcacatatacctta 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTCTTATTACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTA 1380

Qy       1381 ataaccaggattcgggaacaaaggcattctgcgatgttgctctccatggaccattttact 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 ATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACT 1440

Qy       1441 cagcctgccaagctgtgttctacacctttgtttttagacacaagcagcttttgagcggaa 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGAGCGGAA 1500

Qy       1501 acctgaaagaaggtttgcagtatcttcagagtctgaattttgagcggatagtgatgagcc 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 ACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCC 1560

Qy       1561 agctaaatcccctgaagatttgcctgccctcagtggttaacttttttgctgcaatcacaa 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 AGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGCAATCACAA 1620

Qy       1621 ataagtaccagctcgtcttctgctacaccatcattgagaggaacaatcgccagatgctgc 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 ATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAATCGCCAGATGCTGC 1680

Qy       1681 cagtcattaggagtaccgctggaggagactcagtgcagatctgcacaaacccgctggaca 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1681 CAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCCGCTGGACA 1740

Qy   1741 ccttcttcccctttgatccctgtgtgctgaagaggtcaaagaaattcattgatcctattt 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CCTTCTTCCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTT 1800

Qy   1801 atcaggtatgggaagacatgagtgctgaagagctacaggagttcaagaaacccatgaaaa 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ATCAGGTATGGGAAGACATGAGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAA 1860

Qy   1861 aggacatagtggaagatgaagatgatgactttctgaaaggcgaagtgccccagaatgata 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 AGGACATAGTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCAGAATGATA 1920

Qy   1921 ccgtgattgggatcacaccaagctcctttgacacgcatttccgaagtccttcaagtagtg 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAGTCCTTCAAGTAGTG 1980

Qy   1981 tgggctccccacccgtgttgtacatgcaacccagtcccctctgacggcagaaatttgtga 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 TGGGCTCCCCACCCGTGTTGTACATGCAACCCAGTCCCCTCTGACGGCAGAAATTTGTGA 2040

Qy   2041 ctgagatgtgacatttgggattccccat 2068
          ||||||||||||||||||||||||||||
Db   2041 CTGAGATGTGACATTTGGGATTCCCCAT 2068


RESULT    2
HSA272050
LOCUS       HSA272050               3756 bp    mRNA    linear   PRI 07-SEP-2000
DEFINITION  Homo sapiens mRNA for transcription initiation factor IA protein
            (TIF-IA gene).
ACCESSION   AJ272050
VERSION     AJ272050.1  GI:10046713
KEYWORDS    TIF-IA; transcription initiation factor.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3756)
  AUTHORS   Bodem,J., Hoffmann-Rohrer,U., Ross,W., Delius,H., Vingron,M. and
            Grummt,I.
  TITLE     Cloning and functional characterization of transcription initiation
            factor TIF-IA, a growth-dependent regulator of ribosomal RNA
            synthesis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3756)
  AUTHORS   Bodem,J.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2000) Bodem J.J., Molekulare Biologie der Zelle
            II, Deutsches Krebsforschungszentrum, INF.280, 69120 Heidelberg,
            GERMANY
FEATURES             Location/Qualifiers
     source          1..3756
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            23..3756
                     /gene="TIF-IA"
     CDS             23..1978
                     /gene="TIF-IA"
                     /function="transcription initiation factor"
                     /codon_start=1
                     /evidence=experimental
                     /product="transcription initiation factor IA protein"
                     /protein_id="CAC07955.1"
                     /db_xref="GI:10046714"
                     /translation="MAAPLLHTRLPGDAAASSSAVKKLGASRTGISNMRALENDFFNS
                     PPRKTVRFGGTVTEVLLKYKKGETNDFELLKNQLLDPDIKDDQIINWLLEFRSSIMYL
                     TKDFEQLISIILRLPWLNRSQTVVEEYLAFLGNLVSAQTVFLRPCLSMIASHFVPPRV
                     IIKEGDVDVSDSDDEDDNLPANFDTCHRALQIIARYVPSTPWFLMPILVEKFPFVRKS
                     ERTLECYVHNLLRISVYFPTLRHEILELIIEKLLKLDVNASRQGIEDAEETATQTCGG
                     TDSTEGLFNMDEDEETEHETKAGPERLDQMVHPVAERLDILMSLVLSYMKDVCYVDGK
                     VDNGKTKDLYRDLINIFDKLLLPTHASCHVQFFMFYLCSFKLGFAEAFLEHLWKKLQD
                     PSNPAIIRQAAGNYIGSFLARAKFIPLITVKSCLDLLVNWLHIYLNNQDSGTKAFCDV
                     ALHGPFYSACQAVFYXFVFRHKQLLSGNLKEGLQYLQSLNFERIVMSQLNPLKICLPS
                     VVNFFAAITNKYQLVFCYTIIERNNRQMLPVIRSTAGGDSVQICTNPLDTFFPFDPCV
                     LKRSKKFIDPIYQVWEDMSAEELQEFKKPMKKDIVEDEDDDFLKGEVPQNDTVIGITP
                     SSFDTHFRSPSSSVGSPPVLYMQPSPL"
     3'UTR           1979..3756
                     /gene="TIF-IA"
                     /evidence=experimental
BASE COUNT     1060 a    692 c    844 g   1156 t      4 others
ORIGIN

  Query Match             92.8%;  Score 1920;  DB 9;  Length 3756;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2020;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     47 gcggtcgcgttagttcggcccaatggcggcaccgctgcttcacacgcgtttgccgggaga 106
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGGTCGCGTTAGTTCGGCCCAATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGA 60

Qy    107 tgcggccgcttcgtcctctgcagttaagaagctgggcgcgtcgaggactgggatttcaaa 166
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGCGGCCGCTTCGTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAA 120

Qy    167 tatgcgtgcattagagaatgactttttcaattctcccccaagaaaaactgttcggtttgg 226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TATGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTTCGGTTTGG 180

Qy    227 tggaactgtgacagaagtcttgctgaagtacaaaaagggtgaaacaaatgactttgagtt 286
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTT 240

Qy    287 gttgaagaaccagctgttagatccagacataaaggatgaccagatcatcaactggctgct 346
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTTGAAGAACCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGCT 300

Qy    347 agaattccgttcttctatcatgtacttgacaaaagactttgagcaacttatcagtattat 406
          .|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGAATTCCGTTCTTCTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTAT 360

Qy    407 attaagattgccttggttgaatagaagtcaaacagtagtggaagagtatttggcttttct 466
          :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ATTAAGATTGCCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCT 420

Qy    467 tggtaatcttgtatcagcacagactgttttcctcagaccgtgtctcagcatgattgcttc 526
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTC 480

Qy    527 ccattttgtgcctccccgagtgatcattaaggaaggcgatgtagatgtttcagattctga 586
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGA 540

Qy    587 tgatgaagatgataatcttcctgcaaattttgacacatgtcacagagccttgcaaataat 646
          ||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||
Db    541 TGATGAAGATGATAATCTTCCTGCAAATTTTGACACATGTCACAGAGCCTTGCAAATAAT 600

Qy    647 agcaagatatgtaccatcgacaccgtggtttctcatgccaatactggtggaaaaatttcc 706
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCC 660

Qy    707 atttgttcgaaaatcagagagaacactggaatgttacgttcataacttactaaggattag 766
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATTTGTTCGAAAATCAGAGAGAACACTGGAATGTTACGTTCATAACTTACTAAGGATTAG 720

Qy    767 tgtatattttccaaccttgaggcatgaaattctggagcttattattgaaaaactactcaa 826
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TGTATATTTTCCAACCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAAACTACTCAA 780

Qy    827 gttggatgtgaatgcatcccggcagggtattgaagatgctgaagaaacagcaactcaaac 886
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GTTGGATGTGAATGCATCCCGGCAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAAC 840

Qy    887 ttgtggtgggacagattccacggaaggattgtttaatatggatgaagatgaagaaactga 946
```

```
Db    841 TTGTGGTGGGACAGATTCCACGGAAGGATTGTTTAATATGGATGAAGATGAAGAAACTGA 900

Qy    947 acatqaaacaaaqqctqqtcctqaacqqctccqaccaqatqqtqcatcctqtaqccqaqcq 1006
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ACATGAAACAAAGGCTGGTCCTGAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCG 960

Qy   1007 cctqqacatcctqatqtctttqqtttqtcctacatqaaqqatqtctqctatqtaqatqq 1066
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CCTGGCACATCCTGATGTCTTTGGTTTGTCCTACATGAAGGATGTCTGCTATGTAGATGC 1020

Qy   1067 taaqtttqataacqqcaaaacaaaqqatctatatcqcqacctqataaacatctttqacaa 1126
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TAAGTTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATCTTTGACAA 1080

Qy   1127 actcctqtttqccccaccatqcctcctqccatqtacaqttttcatqttttacctctqtaq 1186
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 ACTCCTGTTGCCCACCCATGCCTCCTGCCATGTACAGTTTTCATGTTTTACCTCTGTAG 1140

Qy   1187 tttcaaattqqqattcqcaqaqqcatttttqqaaacatctctqqaaaaaattqcaqqaccc 1246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TTTCAAATTGGGATTCGCAGAGGCATTTTTGGAACATCTCTGGAAAAATTGCAGGACCC 1200

Qy   1247 aaqtaatcctqccatcatcaqqcaqqctqctqqaaattatattqqaaqctttttqqccaaq 1306
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 AAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAG 1260

Qy   1307 aqctaaatttattcctcttattactqtaaaatcatqcctaqatctttqtqttaactqqct 1366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 AGCTAAATTTATTCCTCTTATTACTGTAAAATCATGCCTAGATCTTTGTGTTAACTGGCT 1320

Qy   1367 qcacatatacctaataaccaqqattcqqqaacaaaqqcattctqcqatqttqctcctcca 1426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 GCACATATACCTAATAACCAGGATTCGGCAACAAAGGCATTCTGCGATGTTGCTCTCCA 1380

Qy   1427 tqqaccatttactcaqcctqccaaqctqtattctacacctttqttttaqacacaaqca 1486
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TGGACCATTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCA 1440

Qy   1487 qctttqaqcqqaaaacctqaaaqaaqctttqcaqtatcttcaqaqtctqaattttqaqcq 1546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GCTTTGAGCGGAAACCTGAAAGAAGCTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCG 1500

Qy   1547 qataatqatqaqcccaqctaaatcccctqaaqatttqcctqcccctcaqtqqttaacttttt 1606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CATAGTGATGAGCCCAGCTAAATCCCCTGAACATTTGCCTGCCCTCAGTGGTTAACTTTTT 1560

Qy   1607 tqctqcaatcacaaataaqtaccaqctcqtcttctqctacaccatcattqaqaqqaacaa 1666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TGCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAA 1620

Qy   1667 tcqccaqatqctqccaqtcattaqqaqtaccqctqqaqqaqactcaqtqcaqatctqcac 1726
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TCGCCAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCAC 1680

Qy   1727 aaaccccqctqqacacctttcttccccttttqatcccctqtqtqctqaaqaqqtcaaaqaaatt 1786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 AAACCCGCTGGACACCTTTCTTCCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATT 1740

Qy   1787 cattqatcctatttatcaqqtatqqqaaqacatqaqtqctqaaqaqctacaqqaqttcaa 1846
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CATTGATCCTATTTATCAGGTGTGGGAAGACATGAGTGCTGAAGAGCTACAGGAGTTCAA 1800

Qy   1847 qaaaccctatqaaaaqqacataqtqqaaqatqaaqatqatqacttttctqaaaqqcqaaqt 1906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 GAAACCCATGAAAAGGACATAGTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGT 1860

Qy   1907 qcccctqaatqatacccqtqattqqqatcacaccaaqctcctttqacacqcatttccqaaq 1966
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GCCCCAGAATGATACCGTGATTGGCATCACACCAAGCTCCTTTGACACGCATTTCCGAAG 1920

Qy   1967 tcctcaaqtaqtqtqqqctccccaccqtqttqtacatqcaacccaqtcccctctqacq 2026
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 TCCTTCAAGTAGTGTGGGCTCCCCACCCGTGTTGTACATGCAACCCAGTCCCCTCTGACG 1980

Qy   2027 qcaqaaattqtqactqaqatqtqacatttqqqattcccccat 2068
          ||||||||||||||||||||||||||||||||||||||||||
Db   1981 GCAGAAATTGTGACTGAGATGTGACATTTGGGATTCCCCAT 2022


RESULT    4
AX036047
LOCUS       AX036047                2040 bp    DNA     linear   PAT 15-NOV-2000
DEFINITION  Sequence 1 from Patent WO0055316.
ACCESSION   AX036047
VERSION     AX036047.1  GI:11191586
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2040)
  AUTHORS   Grummt,I. and Vingron,M.
  TITLE     Rna polymerase i transcription factor tif-ia
  JOURNAL   Patent: WO 0055316-A 1 21-SEP-2000;
            DEUTSCHES KREBSFORSCH (DE) ; GRUMMT INGRID (DE) ; VINGRON MARTIN
            (DE)
FEATURES             Location/Qualifiers
     source          1. .2040
                     /organism "Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             40. .1995
                     /note "unnamed protein product"
                     /codon_start 1
                     /protein_id="CAC16268.1"
                     /db xref "GI:11191587"
                     /translation="MAAPLLETRLPGDAAASSSAVKKLGASRTGISNMRALENDFFNS
                     PPRKTVRFGGTVTEVLLKYKKGEFNDFELLKNQLLDPDIKDDQIINWLLEFRSSIMYL
                     TKDFEQLISILRLPWLNRSQTVVEEYLAFLGNLVSAQTVFLRPCLSMIASHFVEPRV
                     IIKEGDVDVSDSDDEDDNLPANFDTCHRALQTIARYVPSTPWFLMPILVEKFPEVRKS
                     EFTLECYVHXLLRISVYFPTLRHEILELIIEKLLKLDVNASRQGIEDAEETATQTCGG
                     TDSTEGLFNMDEDEETEHETKAGPERLDQMVHPVAERLDILLMSLVLSYMKDVCYVDGK
                     VDNGKTKDLYRDLINIFDKLLLPTHASCHVQFFMFYLCSFKLGFAEAFLEHLWKKLQD
                     PSNPAIIRQAAGNYIGSFLARAKFIPLITVKSCLDLLVNWLHIYLNNQDSGTKAFCDV
                     ALHGPFYSACQAVFYXFVFRHKQLLSGNLKEGLQYLQSLNFERIVMSQLNPLKICLPS
                     VVNFFAAITNKYQLVFCYTTERNNRQMLPVIRSTAGGDSVQICTNPLDTFFPFDFCV
                     LKRSKKFIDPIYQVWEDMSAEELQEFKKPMKKDIVEDEDDDFLKGEVPQNDTVIGITP
                     SSFDTHFRSPSSSVGSPPVLYMQPSPL"
BASE COUNT       563 a    433 c    474 g    568 t      2 others
ORIGIN


  Query Match             90.4%;  Score 1869;  DB 6;  Length 2040;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2019;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      47 aeqqtcqcqttaqttcqqcccaatqqcqqcqcaccqctqcttcacacqcqtttqccqqqaqa 106
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18 ACGGTCGCGTTAGTTCGGCCCAATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGA 77

Qy     107 tqcqqcccqcttcqtcctctqcaqttaaqaaqctqqqcqcqtcqaqqactqqqatttcaaa 166
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 TGCGGCCGCTTCGTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAA 137

Qy     167 tatqcqtqcattajaqaatqactttttcaattctcccccaaqaaaaactqttcqqtttqq 226
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1-8 TATGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTTCGGTTTGG 197

Qy     227 tqqaactqtqacaqaaqtcttqctqaaqtacaaaaaqqqtqaaacaaatqactttqaqtt 286
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     198 TGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTT 257

Qy     287 qttqaaqaacqaqctqttaqatccaqacataaaqqatqaccaqatcatcaactqqctqct 346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     258 GTTGAAGAACCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGCT 317
```

```
Qy    347 agaattccgttcttctatcatgtacttgacaaaagactttgagcaacttatcagtattat 406
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    318 AGAATTCCGTTCTTCTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTAT 377

Qy    407 attaagattgccttggttgaatagaagtcaaacagtagtggaagagtatttggcttttct 466
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    378 ATTAAGATTGCCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCT 437

Qy    467 tggtaatcttgtatcagcacagactgttttcctcagaccgtgtctcagcatgattgcttc 526
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    438 TGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTC 497

Qy    527 ccattttgtgcctccccgagtgatcattaaggaaggcgatgtagatgtttcagattctga 586
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    498 CCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGA 557

Qy    587 tgatgaagatgataatcttcctgcaaattttgacacatgtcacagaggccttgcaaataat 646
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    558 TGATGAAGATGATAATCTTCCTGCAAATTTTGACACATGTCACAGAGCCTTGCAAATAAT 617

Qy    647 agcaagatatgtaccatcgacaccgtggtttctcatgccaatactggtggaaaaatttcc 706
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    618 AGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCC 677

Qy    707 atttgttcgaaaatcagagagaacactggaatgttacgttcataacttactaaggattag 766
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    678 ATTTGTTCGAAAATCAGAGAGAACACTGGAATGTTACGTTCATNACTTACTAAGGATTAG 737

Qy    767 tgtatatttttccaaccttgaggcatgaaattctggagcttattattgaaaaactactcaa 826
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    738 TGTATATTTTCCAACCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAAACTACTCAA 797

Qy    827 gttggatgtgaatgcatcccggcagggtattgaagatgctgaagaaacagcaactcaaac 886
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    798 GTTGGATGTGAATGCATCCCGGCAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAAC 857

Qy    887 ttgtggtgggacagattccacggaaggattgtttaatatggatgaaagatgaagaaactga 946
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    858 TTGTGGTGGGACAGATTCCACGGAAGGATTGTTTAATATGGATGAAGATGAAGAAACTGA 917

Qy    947 acatgaaacaaaggctggtcctgaacggctcgaccagatggtgcatcctgtagccgagcg 1006
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    918 ACATGAAACAAAGGCTGGTCCTGAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCG 977

Qy   1007 cctggacatcctgatgtctttggttttgtcctacatgaaggatgtctgctatgtagatgg 1066
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    978 CCTGGACATCCTGATGTCTTTGGTTTTGTCCTACATGAAGGATGTCTGCTATGTAGATGG 1037

Qy   1067 taaggttgataacggcaaaacaaaggatctatatcgcgacctgataaacatctttgacaa 1126
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1038 TAAGGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATCTTTGACAA 1097

Qy   1127 actcctgttgcccacccatgcctcctgccatgtacagtttttcatgtttacctctgtag 1186
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1098 ACTCCTGTTGCCCACCCATGCCTCCTGCCATGTACAGTTTTTCATGTTTACCTCTGTAG 1157

Qy   1187 tttcaaattgggattcgcagaggcatttttggaacatctctggaaaaaattgcaggaccc 1246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1158 TTTCAAATTGGGATTCGCAGAGGCATTTTTGGAACATCTCTGGAAAAAATTGCAGGACCC 1217

Qy   1247 aagtaatcctgccatcatcaggcaggctgctgaaaattatattggaagctttttggcaag 1306
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1218 AAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAG 1277

Qy   1307 agctaaatttattcctcttattactgtaaaatcatgcctagatcttttggttaactggct 1366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1278 AGCTAAATTTATTCCTCTTATTACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCT 1337

Qy   1367 gcacatatacccttaataaccaggattcgggaacaaaggcattctgcgatgttgctctcca 1426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1338 GCACATATACCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCA 1397

Qy   1427 tggaccattttactcagcctgccaagctgtgttctacaccttttgtttttaqacacaaqca 1486
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1398 TGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACNCCTTTGTTTTTAGACACAAGCA 1457

Qy   1487 gcttttgaagcggaaacctgaaagaaggtttgcagtatcttcagagtctgaatttgaagcg 1546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1458 GCTTTTGAGCGGAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTGAGCG 1517

Qy   1547 gataagtgatgagccagctaaatcccctgaaagatttgcctgccctcagtggttaacttttt 1606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1518 GATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTT 1577

Qy   1607 tgctgcaatcacaaataagtaccagctcgtcttctgctacaccatcattgagaggaacaa 1666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1578 TGCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAA 1637

Qy   1667 tcgccagatgctgccagtcattaggagtaccgctggaggagactcagtgcagatctgcac 1726
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1638 TCGCCAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCAC 1697

Qy   1727 aaaccccgctggacaccttcttccccttttgatccctgtgtgctgaaggagtcaaagaaatt 1786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1698 AAACCCGCTGGACACCTTCTTCCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATT 1757

Qy   1787 cattgatcctatttatcaggtatgggaagacatgagtgctgaagagctacaggagttcaa 1846
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1758 CATTGATCCTATTTATCAGGTGTGGGAAGACATGAGTGCTGAAGAGCTACAGGAGTTCAA 1817

Qy   1847 gaaacccatgaaaaaggacatagtggaagatgaagatgatgacttttctgaaaggcgaagt 1906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1818 GAAACCCATGAAAAAGGACATAGTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGT 1877

Qy   1907 gccccagaatgataccgtgattgggatcacaccaagctccttgacacgcatttccgaag 1966
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1878 GCCCCAGAATGATACCGTGATTGGGATCACACCAAGCTCCTTGACACGCATTTCCGAAG 1937

Qy   1967 tccttcaagtagtgtgggctcccaccgtgttgtacatgcaacccagtcccctctgacg 2026
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1938 TCCTTCAAGTAGTGTGGGCTCCCACCGTGTTGTACATGCAACCCAGTCCCCTCTGACG 1997

Qy   2027 gcagaaatttgtgactgagatgtgacatttgggattccccat 2068
          |||||||||||||||||||||||||||||||||||||||||
Db   1998 GCAGAAATTTGTGACTGAGATGTGACATTTGGGATTCCCCAT 2049
```

```
RESULT   4
AK055742
LOCUS       AK055742               2493 bp    mRNA    linear   PRI 31-OCT 2001
DEFINITION  Homo sapiens cDNA FLJ31180 fis, clone KIDNE2000266.
ACCESSION   AK055742
VERSION     AK055742.1  GI:16550546
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens kidney cDNA to mRNA, clone_lib:KIDNE2
            clone:KIDNE2000266.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,
            Ono,Y., Hotuta,T., Hiraoka,S., Murakawa,K., Takiguchi,S.,
            Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,
            Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,
            Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai Hio,Y.,
            Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y.,
            Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2493)
  AUTHORS   Isogai,T., Otsuki,T. and Sugiyama,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT 2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
```

```
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: RAB and
            HRI.
FEATURES             Location/Qualifiers
     source          1..2493
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "KIDNE2000266"
                     /tissue_type "kidney"
                     /clone_lib "KIDNE2"
                     /note "cloning vector: pME18SFL3"
BASE COUNT      554 a    700 c    761 g    478 t
ORIGIN


  Query Match             14.9%;  Score 808;  DB 9;  Length 2493;
  Best Local Similarity   99.5%;  Pred. No. 9.1e-151;
  Matches  408;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    1345 ccaaqtaatcctqccatcatcaqqcaqqctqctqqaaattatattqqaaqctttttqqca 1404
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     760 CCAAGTAATCCTGCCATCATCAGCCAGGCTGCTGGAAATTATATTGCAAGCTTTTTGGCA 819

Qy    1405 aqaqctaaatttattcctcttattactqtaaaatcatqcctaqatcttttqqttaactqq 1464
           ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     820 AGAGCTAAATTTATTCCTCTTATTACTGTAAAACCATGCCTAGATCTTTTGGTTAACTGG 879

Qy    1465 ctqcacatataccttaataaccaqqattcqqqaacaaaqqcattctqcqatqttqctctc 1524
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     880 CTGCACATATACCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCAATGTTGCTCTC 939

Qy    1525 catqqaccatttactcaqcctqccaaqctqtqttctacacctttqttttaqacacaaq 1584
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     940 CATGGACCATTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTCTTTTTAGACACAAG 999

Qy    1585 caqctttqaqcqqaaacctqaaaqaaqqtttqcaqtatatcttcaqaqtctqaattttqaq 1644
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1000 CAGCTTTTGAGCGGAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAG 1059

Qy    1645 cqqataqtqatqaqccaqctaaatcccctqaaqatttqcctqccctcaqtqqttaacttt 1704
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1060 CGGATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTT 1119

Qy    1705 tttqctqcaatcacaaataaqtaccaqctcqtctttctqctacaccatcat 1754
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1120 TTTGCTGCAATCACAAATAAGTACCAGCTCGTCTTTCTGCTACACCATCAT 1169


RESULT   5
BC009198
LOCUS       BC009198                 1177 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, Similar to RNA polymerase I transcription factor
            RRN3, clone MGC:15321 IMAGE:4678732, mRNA, complete cds.
ACCESSION   BC009198
VERSION     BC009198.1  GI:14327947
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1177)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 24 Row: k Column: 14
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4455402.
FEATURES             Location/Qualifiers
     source          1..1177
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "MGC:15321 IMAGE:4678732"
                     /tissue_type "Lymph, Burkitt lymphoma"
                     /clone_lib "NIH MGC_8"
                     /lab_host="DH10B R"
                     /note "Vector: pOTB7"
     CDS             316..636
                     /codon_start=1
                     /product "Similar to RNA polymerase I transcription factor
                     RRN3"
                     /protein_id="AAH09198.1"
                     /db_xref "GI:14327948"
                     /translation "MRALENDFFNSPPRKTVRFGGTVTEVLLKYKKGFTNDFELLKNQ
                     LLDPDIKDDQIINWLLEFRSSVMYLTKDFEQLISIHLRLPWLNRSQTVVEEYLAFLGN
                     LVSA"
BASE COUNT      339 a    222 c    253 g    363 t
ORIGIN


  Query Match             13.7%;  Score 283;  DB 9;  Length 1177;
  Best Local Similarity   99.3%;  Pred. No. 1.5e-137;
  Matches  433;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy     189 ttttcaattctccccccaaqaaaaactqttcqqtttqqtqqaactqtqacaqaaqtcttq 248
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     337 TTTTCAATTCTCCCCCAAGAAAAACTGTTCGGTTTGGTGGAACTGTGACAGAAGTCTTG 396

Qy     249 ctqaaqtacaaaaaaqqtqaaacaaatqacttqaqttqttqaaqaaccaqctqttaqat 308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     397 CTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGAT 456

Qy     309 ccaqacataaaqqatqaccaqatcatcaactqqctqctaqaattccqttcttctatcatq 368
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     457 CCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTGTCATG 516

Qy     369 tacttqacaaaaqacttqaqcaacttatcaqtattatattaaqattqccttqqttqaat 428
           |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db     517 TACTTGACAAAAGACTTGAGCAACTTATCAGTATTATATTGAGATTGCCTTGGTTGAAT 576

Qy     429 aqaaqtcaaacaqtaqtqqaaqaqtatttqqcttttcttqqtaatcttqtatcaqcacaq 488
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     577 AGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTGGTAATCTTGTATCAGCATAG 636

Qy     489 actqttttcctcaqaccqtqtctcaqcatqattqcttcccattttqtqcctccccqaqtq 548
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     637 ACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGTGCCTCCCCGAGTG 696

Qy     549 atcattaaqqaaqqcqatqtaqatqtttcaqattctqatqatqaaqatqataatcttcct 608
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     697 ATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCT 756

Qy     609 qcaaattttqacacat 624
           ||||||||||||||||
```

```
Db     757 GCAAATTTTGACACAT 772


RESULT   6
AX274907
LOCUS       AX274907               1461 bp    DNA     linear   PAT 29-OCT-2001
DEFINITION  Sequence 172 from Patent WO0172777.
ACCESSION   AX274907
VERSION     AX274907.1  GI:16547539
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Hillman,J.L., Baughn,M.R., Yue,H., Lal,P., Lu,D.A., Patterson,C.,
            Azimzai,Y., Bandman,O., Tang,Y.T., Mathur,P., Shah,P., Au-Young,J.
            and Reddy,R.
  TITLE     Transcription factors
  JOURNAL   Patent: WO 0172777-A 172 04-OCT-2001;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1461
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 2646274CB1"
BASE COUNT      413 a    335 c    298 g    415 t
ORIGIN


  Query Match             11.2%;  Score 232;  DB 6;  Length 1461;
  Best Local Similarity   99.6%;  Pred. No. 1.2e-110;
  Matches  282;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    1339 catgcctagatcttttggttaactggctgcacatataccttaataaccaggattcgggaa 1398
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     681 CATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCAGGATTCGGGAA 740

Qy    1399 caaaggcattctgcgatgttgctctccatggaccatttttactcagcctgccaagctgtgt 1458
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     741 CAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGT 800

Qy    1459 tctacacctttgttttttagacacaagcagctttttgaggcggaaacctgaaaagaaggtttgc 1518
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     801 TCTACACCTTTGTTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGTTTGC 860

Qy    1519 agtatcttcagagtctgaatttttgaagcggatagtgatgagccagctaaatccccctgaaga 1578
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     861 AGTATCCTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAATCCCCTGAAGA 920

Qy    1579 tttgcctgcccctcagtggttaacttttttgctgcaatcacaaa 1621
           |||||||||||||||||||||||||||||||||||||||||||||
Db     921 TTTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGCAATCACAAA 963


RESULT   7
BC006441
LOCUS       BC006441               2627 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, Similar to RNA polymerase I transcription factor
            RRN3, clone MGC:13169 IMAGE:3351791, mRNA, complete cds.
ACCESSION   BC006441
VERSION     BC006441.1  GI:13623642
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2627)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapos-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 18 Row: d Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.
FEATURES             Location/Qualifiers
     source          1. .2627
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:13169 IMAGE:3351791"
                     /tissue_type="Eye, retinoblastoma"
                     /clone_lib="NIH_MGC_16"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             1698. .2156
                     /codon_start=1
                     /product="Similar to RNA polymerase I transcription factor
                     RRN3"
                     /protein_id="AAH06441.1"
                     /db_xref="GI:13623643"
                     /translation="MGFAEAFLEHLWKNLQDPSNPAIIRQAACNYIGSFLARAKFISL
                     ITVKPCLDLLVNWLHIYLNNQDSGTKAFCDVALHGPFYSACQAVFYTFVFKHKQLLSG
                     NLKEGLQYPQSLNFERIVMSQLNPLKICLPSVVNFFAAIIKMKTCGYGWW"
BASE COUNT      837 a    492 c    525 g    773 t
ORIGIN


  Query Match             11.2%;  Score 232;  DB 9;  Length 2627;
  Best Local Similarity   99.6%;  Pred. No. 1.2e-110;
  Matches  282;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    1339 catgcctagatcttttggttaactggctgcacatatacccttaataaccaggattcgggaa 1398
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1843 CATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCAGGATTCGGGAA 1902

Qy    1399 caaaggcattctgcgatgttgctctccatggaccatttttactcagcctgccaagctgtgt 1458
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1903 CAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGT 1962

Qy    1459 tctacacctttgttttttagacacaagcagctttttgaggcggaaacctgaaaagaaggtttgc 1518
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1963 TCTACACCTTTGTTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGTTTGC 2022

Qy    1519 agtatcttcagagtctgaatttttgaagcggatagtgatgagccagctaaatccccctgaaga 1578
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2023 AGTATCCTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAATCCCCTGAAGA 2082

Qy    1579 tttgcctgcccctcagtggttaacttttttgctgcaatcacaaa 1621
           |||||||||||||||||||||||||||||||||||||||||||||
Db    2083 TTTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGCAATCACAAA 2125


RESULT   8
```

```
AC040158
LOCUS       AC040158               129473 bp    DNA     linear   HTG 03-SEP-2000
DEFINITION  Homo sapiens chromosome 16 clone CTA-133B4, WORKING DRAFT SEQUENCE,
            14 unordered pieces.
ACCESSION   AC040158
VERSION     AC040158.3  GI:9965546
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 129473)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 129473)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (11 APR 2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     on Sep 3, 2000 this sequence version replaced gi:9090652.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------

            Project Information
            Center Project Name: 0
            Center clone name: CIT978SKA_133B4
            ----

            Summary Statistics
            Consensus quality: 90046 bases at least Q40
            Consensus quality: 99266 bases at least Q30
            Consensus quality: 104098 bases at least Q20
            Estimated insert size: 125000; agarose fp estimation
            Estimated insert size: 128174; sum-of-contigs estimation
            Quality coverage:  4.51 in Q20 bases; agarose-fp estimation
            Quality coverage:  4.4 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1718: contig of 1718 bp in length
            *     1719     1818: gap of unknown length
            *     1819     3832: contig of 2014 bp in length
            *     3833     3932: gap of unknown length
            *     3933     5567: contig of 1635 bp in length
            *     5568     5667: gap of unknown length
            *     5668     7520: contig of 1853 bp in length
            *     7521     7620: gap of unknown length
            *     7621     9146: contig of 1526 bp in length
            *     9147     9246: gap of unknown length
            *     9247    11144: contig of 1898 bp in length
            *    11145    11244: gap of unknown length
            *    11245    12746: contig of 1502 bp in length
            *    12747    12846: gap of unknown length
            *    12847    14568: contig of 1722 bp in length
            *    14569    14668: gap of unknown length
            *    14669    19396: contig of 4728 bp in length
            *    19397    19496: gap of unknown length
            *    19497    25634: contig of 6138 bp in length
            *    25635    25734: gap of unknown length
            *    25735    36780: contig of 11046 bp in length
            *    36781    36880: gap of unknown length
            *    36881    50223: contig of 13343 bp in length
            *    50224    50423: gap of unknown length
            *    50424    71412: contig of 20989 bp in length
            *    71413    71412: gap of unknown length
            *    71413   129473: contig of 58061 bp in length.

FEATURES             Location/Qualifiers
     source          1..129473
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "16"
                     /clone "CTA-133B4"
                     /clone_lib "CalTech human BAC library A"
BASE COUNT    30715 a   30506 c   31241 g   35543 t    1468 others
ORIGIN


  Query Match           10.1%;  Score 208;  DB 2;  Length 129473;
  Best Local Similarity   100.0%;  Pred. No. 4e-98;
  Matches  208;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1861 aqqacatagtqqaaqatqaaqatqatqacttttctqaaaqqcqaaqtqccccaqaatgata 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  123987 ACGACATAGTCGAAGATGAAGATGATGACTTTCTCGAAAGCCGAAGTGCCCCAGAATGATA 124046

Qy    1921 ccqtqattqqqatcacaccaaqctccttqacacqcatttccqaaqtccttcaaqtaqtq 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124047 CCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAGTCCTTCAAGTAGTG 124106

Qy    1981 tqqqctcccqcacccqtqttqtacatqcaacccaqtccccctctqacqqcaaaaatttqtqa 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124107 TGGGCTCCCCACCCGTGTTGTACATGCAACCCAGTCCCCTCTGACGGCAGAAATTTGTGA 124166

Qy    2041 ctqaqatqtqacatttqqqattccccat 2068
           |||||||||||||||||||||||||||||
Db  124167 CTGACATGTGACATTTGGGATTCCCCAT 124194


RESULT   9
AC092147/C
LOCUS       AC092147               143900 bp    DNA     linear   HTG 22-JUN-2001
DEFINITION  Homo sapiens chromosome 16 clone RPCI-11_44H15, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.
ACCESSION   AC092147
VERSION     AC092147.1  GI:14522983
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 143900)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 143900)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 538644
            Center clone name: RPCI-11_44H15
            ----
            Summary Statistics
            Consensus quality: 124071 bases at least Q40
            Consensus quality: 136140 bases at least Q30
            Consensus quality: 137741 bases at least Q20
            Estimated insert size: 211000; agarose-fp estimation
            Estimated insert size: 142200; sum-of-contigs estimation
            Quality coverage:  4.0 in Q20 bases; agarose fp estimation
            Quality coverage:  5.94 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
```

```
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1033: contig of 1033 bp in length
            *     1034     1133: gap of unknown length
            *     1134     2547: contig of 1414 bp in length
            *     2548     2647: gap of unknown length
            *     2648     4205: contig of 1558 bp in length
            *     4206     4305: gap of unknown length
            *     4306     6564: contig of 2259 bp in length
            *     6565     6664: gap of unknown length
            *     6665     9065: contig of 2401 bp in length
            *     9066     9165: gap of unknown length
            *     9166    12059: contig of 2894 bp in length
            *    12060    12159: gap of unknown length
            *    12160    19295: contig of 7136 bp in length
            *    19296    19395: gap of unknown length
            *    19396    25668: contig of 6273 bp in length
            *    25669    25768: gap of unknown length
            *    25769    31013: contig of 5245 bp in length
            *    31014    31113: gap of unknown length
            *    31114    36555: contig of 5442 bp in length
            *    36556    36655: gap of unknown length
            *    36656    45338: contig of 8683 bp in length
            *    45339    45438: gap of unknown length
            *    45439    52918: contig of 7480 bp in length
            *    52919    53018: gap of unknown length
            *    53019    65720: contig of 12702 bp in length
            *    65721    65820: gap of unknown length
            *    65821    76109: contig of 10289 bp in length
            *    76110    76209: gap of unknown length
            *    76210    85439: contig of 9230 bp in length
            *    85440    85539: gap of unknown length
            *    85540   101620: contig of 16081 bp in length
            *   101621   101720: gap of unknown length
            *   101721   122163: contig of 20443 bp in length
            *   122164   122263: gap of unknown length
            *   122264   143900: contig of 21637 bp in length.
FEATURES             Location/Qualifiers
     source          1. .143900
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-344H15"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT    37791 a   34128 c   34320 g   35886 t    1775 others
ORIGIN


  Query Match             10.1%;  Score 208;  DB 2;  Length 143900;
  Best Local Similarity   100.0%;  Pred. No. 4e-98;
  Matches  208;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1861 agqacataqtqqaaqatqaaqatqatqacttttctqaaaqqcqaaqtqccccaqaatqata 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   27338 AGGACATAGTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCAGAATGATA 27279

Qy    1921 ccgtgattgggatcacaccaaqctcctttgacacgcatttccgaagtcctttcaagtagtg 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   27278 CCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAGTCCTTCAAGTAGTG 27219

Qy    1981 tqqqctcccaacccgtgttgtacatqcaacccaqtcccctctqacqqcaqaaatttgtqa 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   27218 TGGGCTCCCCACCCGTGTTGTACATGCAACCCAGTCCCCTCTGACGGCAGAAATTTGTGA 27159

Qy    2041 ctgagatgtgacatttqqqattccccat 2068
           |||||||||||||||||||||||||||||
Db   27158 CTGAGATGTGACATTTGGGATTCCCCAT 27131


RESULT  10
AC007615
LOCUS       AC007615               175691 bp    DNA     linear   HTG 03-JUL-2001
DEFINITION  Homo sapiens chromosome 16 clone RP11-528K16, WORKING DRAFT
            SEQUENCE, 1 ordered pieces.
ACCESSION   AC007615
VERSION     AC007615.6  GI:14589428
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175691)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175691)
  AUTHORS   Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
            Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
            Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
            Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
            McMurry,K., Han,C. and Deaven,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint
            Genome Institute, Los Alamos National Laboratory, MS M888, Los
            Alamos, NM 87545, USA
COMMENT     On Jul 3, 2001 this sequence version replaced gi:13928651.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
            ----------.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1   175691: contig of 175691 bp in length.
FEATURES             Location/Qualifiers
     source          1. .175691
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-528K16"
BASE COUNT    42392 a   42125 c   45277 g   45896 t       1 others
ORIGIN


  Query Match             9.0%;  Score 187;  DB 2;  Length 175691;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-87;
  Matches  187;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1328 tactgtaaaatcatgcctagatcttttggttaactggctgcacatataccttaataacca 1387
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31066 TACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCA 31065
```

```
Qy    1388 qgattcqqqaacaaaqqcattctqcqatqttqctctccatqqaccatttttactcaqccctq 1447
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41066 GGATTCGGCAACAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTG 41125

Qy    1448 ccaaqctqtqttctacacctttqttttaqacacaaqcaqctttqaqcqqaaaacctqaa 1507
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41126 CCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAA 41185

Qy    1508 aqaaqt 1514
           ||||||
Db   41186 AGAAGT 41192


RESULT  11
AC009140
LOCUS       AC009140              191496 bp    DNA     linear   HTG 21-JUN-2000
DEFINITION  Homo sapiens chromosome 16 clone RP11 501P17, WORKING DRAFT
            SEQUENCE, 43 unordered pieces.
ACCESSION   AC009140
VERSION     AC009140.6  GI:8575965
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191496)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191496)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03 AUG 1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jun 21, 2000 this sequence version replaced gi:7689948.
                  Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov

            Project Information
            Center Project Name: 599126
            Center clone name: RPCI-11 501P17

            Summary Statistics
            Consensus quality: 158339 bases at least Q40
            Consensus quality: 176831 bases at least Q30
            Consensus quality: 180633 bases at least Q20
            Estimated insert size: 201840; agarose fp estimation
            Estimated insert size: 187296; sum-of-contigs estimation
            Quality coverage:  3.4 in Q20 bases; agarose-fp estimation
            Quality coverage:  3.67 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 43 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1071: contig of 1071 bp in length
            *     1072     1171: gap of unknown length
            *     1172     2358: contig of 1187 bp in length
            *     2359     2458: gap of unknown length
            *     2459     3680: contig of 1222 bp in length
            *     3681     3780: gap of unknown length
            *     3781     5278: contig of 1498 bp in length
            *     5279     5378: gap of unknown length
            *     5379     6522: contig of 1144 bp in length
            *     6523     6622: gap of unknown length
            *     6623     8589: contig of 1967 bp in length
            *     8590     8689: gap of unknown length
            *     8690    10266: contig of 1577 bp in length
            *    10267    10366: gap of unknown length
            *    10367    11714: contig of 1348 bp in length
            *    11715    11814: gap of unknown length
            *    11815    13443: contig of 1629 bp in length
            *    13444    13543: gap of unknown length
            *    13544    14645: contig of 1102 bp in length
            *    14646    14745: gap of unknown length
            *    14746    15751: contig of 1006 bp in length
            *    15752    15851: gap of unknown length
            *    15852    16868: contig of 1017 bp in length
            *    16869    16968: gap of unknown length
            *    16969    19037: contig of 2069 bp in length
            *    19038    19137: gap of unknown length
            *    19138    20794: contig of 1657 bp in length
            *    20795    20894: gap of unknown length
            *    20895    22876: contig of 1982 bp in length
            *    22877    22976: gap of unknown length
            *    22977    24751: contig of 1775 bp in length
            *    24752    24851: gap of unknown length
            *    24852    27481: contig of 2630 bp in length
            *    27482    27481: gap of unknown length
            *    27482    30339: contig of 2858 bp in length
            *    30440    30439: gap of unknown length
            *    30440    32397: contig of 1958 bp in length
            *    32398    32497: gap of unknown length
            *    32498    34880: contig of 2383 bp in length
            *    34881    34980: gap of unknown length
            *    34981    37578: contig of 2598 bp in length
            *    37579    37678: gap of unknown length
            *    37679    40004: contig of 2326 bp in length
            *    40005    40104: gap of unknown length
            *    40105    43544: contig of 3440 bp in length
            *    43545    43644: gap of unknown length
            *    43645    46429: contig of 2785 bp in length
            *    46430    46529: gap of unknown length
            *    46530    49686: contig of 3157 bp in length
            *    49687    49786: gap of unknown length
            *    49787    52269: contig of 2483 bp in length
            *    52270    52369: gap of unknown length
            *    52370    55195: contig of 2826 bp in length
            *    55196    55295: gap of unknown length
            *    55296    58516: contig of 4221 bp in length
            *    58517    58616: gap of unknown length
            *    58617    61359: contig of 2743 bp in length
            *    61360    61459: gap of unknown length
            *    61460    66151: contig of 4692 bp in length
            *    66152    66251: gap of unknown length
            *    66252    73145: contig of 6894 bp in length
            *    73146    73245: gap of unknown length
            *    73246    78200: contig of 4955 bp in length
            *    78201    78300: gap of unknown length
            *    78301    83186: contig of 4886 bp in length
            *    83187    83286: gap of unknown length
            *    83287    88570: contig of 5284 bp in length
            *    88571    88670: gap of unknown length
            *    88671    95276: contig of 6606 bp in length
            *    95277    95376: gap of unknown length
            *    95377   102234: contig of 6858 bp in length
            *   102235   102334: gap of unknown length
            *   102335   107640: contig of 5296 bp in length
            *   107641   107730: gap of unknown length
            *   107731   120220: contig of 12490 bp in length
            *   120221   120320: gap of unknown length
            *   120321   134451: contig of 14131 bp in length
            *   134452   134551: gap of unknown length
            *   134552   144891: contig of 10340 bp in length
            *   144892   144991: gap of unknown length
            *   144992   159434: contig of 14443 bp in length
            *   159435   159534: gap of unknown length
            *   159535   175863: contig of 16329 bp in length
            *   175864   175963: gap of unknown length
            *   175964   191496: contig of 15533 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1. .191496
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-501P17"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT    47031 a  47510 c  47005 g  45732 t   4218 others
ORIGIN


  Query Match             9.0%;  Score 187;  DB 2;  Length 191496;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-87;
  Matches  187;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1328 tactgtaaaatcatgcctagatcttttggttaactggctgcacatatacattaataacca 1387
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   26623 TACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCA 26682

Qy    1388 gqattcgggaacaaaggcattctgcgatgttgctctccatggaccatttttactcagcctg 1447
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   26683 GGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTG 26742

Qy    1448 ccaaqctgtgttctacaccttttgtttttagacacaaqcaqctttgaqcggaaaacctqaa 1507
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   26743 CCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTGAGCGGAAACCTGAA 26802

Qy    1508 agaaggt 1514
           |||||||
Db   26803 AGAAGGT 26809


RESULT  12
AC009093
LOCUS       AC009093              208008 bp    DNA     linear   HTG 04-SEP-2001
DEFINITION  Homo sapiens chromosome 16 clone RP11-426C22, WORKING DRAFT
            SEQUENCE, 1 ordered pieces.
ACCESSION   AC009093
VERSION     AC009093.8  GI:15426058
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 208008)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 208008)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Sep 4, 2001 this sequence version replaced gi:13786300.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
            ----------.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   208008: contig of 208008 bp in length.
FEATURES             Location/Qualifiers
     source          1. .208008
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-426C22"
BASE COUNT    49489 a  50452 c  53814 g  54252 t      1 others
ORIGIN


  Query Match             9.0%;  Score 187;  DB 2;  Length 208008;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-87;
  Matches  187;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1328 tactgtaaaatcatgcctagatcttttggttaactggctgcacatatacattaataacca 1387
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   28488 TACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCA 28547

Qy    1388 gqattcgggaacaaaggcattctgcgatgttgctctccatggaccatttttactcagcctg 1447
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   28548 GGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTG 28607

Qy    1448 ccaaqctgtgttctacaccttttgtttttagacacaaqcaqctttgaqcggaaaacctqaa 1507
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   28608 CCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTGAGCGGAAACCTGAA 28667

Qy    1508 agaaggt 1514
           |||||||
Db   28668 AGAAGGT 28674


RESULT  13
AC092375/c
LOCUS       AC092375              173166 bp    DNA     linear   PRI 02-NOV-2001
DEFINITION  Homo sapiens chromosome 16 clone RP11-645C24, complete sequence.
ACCESSION   AC092375
VERSION     AC092375.3  GI:16596541
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173166)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173166)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 173166)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Nov 2, 2001 this sequence version replaced gi:15667998.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
```

```
            1 error in 10,000 bp.
            Base by base quality values are not generally visible from the
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.

           Sequence Quality Assessment:
            This entry has been annotated with sequence quality
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality zero.
            Quality levels above 40 are expected to have less than
            1 error in 10,000 bp.
            Base by base quality values are not generally visible from the
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.

FEATURES             Location/Qualifiers
     source          1..173166
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "16"
                     /clone "RP11 645C24"
     misc feature    61420..63375
                     /note "unresolved tandem repeat"
BASE COUNT     51459 a  37672 c  36529 g  47506 t
ORIGIN


  Query Match             8.5%;  Score 176;  DB 9;  Length 173166;
  Best Local Similarity   100.0%;  Pred. No. 3e-81;
  Matches  176;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1339 catgcctagatcttttggttaactggctgcacatataccttaataaccaggattcgggaa 1398
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     26106 CATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCAGGATTCGGGAA 26047

Qy      1399 caaaggcattctgcgatgttgctctccatggaccattttactcagcctgccaagctgtgt 1458
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     26046 CAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGT 25987

Qy      1459 tctacacctttgtttttagacacaagcagcttttgagcggaaacctgaaagaaggt 1514
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25986 TCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGT 25931


RESULT 14
AC017077/c
LOCUS       AC017077              179150 bp    DNA     linear   HTG 25-JUL-2000
DEFINITION  Homo sapiens chromosome 16 clone RP11-449K24, WORKING DRAFT
            SEQUENCE, 19 unordered pieces.
ACCESSION   AC017077
VERSION     AC017077.5  GI:9443443
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179150)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179150)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09 DEC 1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jul 25, 2000 this sequence version replaced gi:8570197.

                          Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ------------ Project Information ------------
```

```
            Center project name: H_NH0449K24
            --------------- Summary Statistics
            Sequencing vector: M13; 83%
            Sequencing vector: plasmid; 17%
            Chemistry: Dye-primer ET; 83% of reads
            Chemistry: Dye-terminator Big Dye; 17% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 167439 bases at least Q40
            Consensus quality: 171134 bases at least Q30
            Consensus quality: 173157 bases at least Q20
            Insert size: 190000; agarose-fp
            Insert size: 177350; sum of contigs
            Quality coverage: 0.00 in Q20 bases; agarose-fp
            Quality coverage: 4.51 in Q20 bases; sum-of-contigs
            ----------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *          1      1329: contig of 1329 bp in length
            *       1330      1429: gap of unknown length
            *       1430      2872: contig of 1443 bp in length
            *       2873      2972: gap of unknown length
            *       2973      5370: contig of 2398 bp in length
            *       5371      5470: gap of unknown length
            *       5471      7617: contig of 2147 bp in length
            *       7618      7717: gap of unknown length
            *       7718      9983: contig of 2266 bp in length
            *       9984     10083: gap of unknown length
            *      10084     13594: contig of 3511 bp in length
            *      13595     13694: gap of unknown length
            *      13695     18473: contig of 4779 bp in length
            *      18474     18573: gap of unknown length
            *      18574     24956: contig of 6383 bp in length
            *      24957     25056: gap of unknown length
            *      25057     34405: contig of 9349 bp in length
            *      34406     34505: gap of unknown length
            *      34506     43064: contig of 8559 bp in length
            *      43065     43164: gap of unknown length
            *      43165     50773: contig of 7609 bp in length
            *      50774     50873: gap of unknown length
            *      50874     60610: contig of 9737 bp in length
            *      60611     60710: gap of unknown length
            *      60711     69701: contig of 8991 bp in length
            *      69702     69801: gap of unknown length
            *      69802     80031: contig of 10230 bp in length
            *      80032     80131: gap of unknown length
            *      80132     93512: contig of 13381 bp in length
            *      93513     93612: gap of unknown length
            *      93613    111039: contig of 17427 bp in length
            *     111040    111139: gap of unknown length
            *     111140    127223: contig of 16084 bp in length
            *     127224    127323: gap of unknown length
            *     127324    149676: contig of 22353 bp in length
            *     149677    149776: gap of unknown length
            *     149777    179150: contig of 29374 bp in length.
FEATURES             Location/Qualifiers
     source          1..179150
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "16"
                     /clone "RP11 449K24"
     misc_feature    1..1329
                     /note "assembly_name:Contig24"
     misc_feature    1430..2872
                     /note "assembly_name:Contig25"
     misc_feature    2973..5370
                     /note "assembly_name:Contig26"
     misc_feature    5471..7617
```

```
                     /note="assembly_name:Contig27"
     misc_feature    7718. .9983
                     /note="assembly_name:Contig28"
     misc_feature    10084. .13594
                     /note="assembly_name:Contig29"
     misc_feature    13695. .18473
                     /note="assembly_name:Contig30"
     misc_feature    18574. .24956
                     /note="assembly_name:Contig31"
     misc_feature    25057. .34405
                     /note="assembly_name:Contig32"
     misc_feature    34506. .43064
                     /note="assembly_name:Contig33"
     misc_feature    43165. .50773
                     /note="assembly_name:Contig34"
     misc_feature    50874. .60610
                     /note="assembly_name:Contig35"
     misc_feature    60711. .69701
                     /note="assembly_name:Contig36"
     misc_feature    69802. .80031
                     /note="assembly_name:Contig37"
     misc_feature    80132. .93512
                     /note="assembly_name:Contig38
                     clone_end:SP6
                     vector_side:left"
     misc_feature    93613. .111039
                     /note="assembly_name:Contig39"
     misc_feature    111140. .127223
                     /note="assembly_name:Contig40"
     misc_feature    127324. .149676
                     /note="assembly_name:Contig41"
     misc_feature    149777. .179150
                     /note="assembly_name:Contig42"
BASE COUNT    51425 a  39220 c  37611 g  49085 t   1809 others
ORIGIN


  Query Match             8.5%;  Score 176;  DB 2;  Length 179150;
  Best Local Similarity   100.0%;  Pred. No. 3e-81;
  Matches  176;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1339 catgcctagatcttttggttaactggctgcacatatacctaataaccaggattcgggaa 1398
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   72665 CATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCAGGATTCGGGAA 72606

Qy    1399 caaaggcattctgcgatgttgctctccatggaccatttttactcagcctgccaagctgtgt 1458
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   72605 CAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGT 72546

Qy    1459 tctacacctttgttttttagacacaagcagctttttgagcggaaaacctgaaagaaggt 1514
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   72545 TCTACACCTTTGTTTTTTAGACACAAGCAGCTTTTGAGCGGAAAACCTGAAAGAAGGT 72490


RESULT  15
HUAF001549
LOCUS       HUAF001549            202004 bp    DNA     linear   PRI 28-JUL-1998
DEFINITION  Human Chromosome 16 BAC clone CIT987SK-A-270G1, complete sequence.
ACCESSION   AF001549
VERSION     AF001549.1  GI:3355302
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 202004)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1997) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   2  (bases 1 to 202004)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., La Bombard,M., Kim,U.J. and
            Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-1997) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 202004)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jul 30, 1998 this sequence version replaced gi:2340964.
            BAC clone CIT987SK-270G1 is located in chromosome 16. Genes were
            identified by a combination of five methods: XGRAIL (available by
            anonymous ftp from arthur.epm.ornl.gov), Genefinder (available by
            anonymous ftp from colin@u.washington.edu), GENSCAN (available
            using the e-mail server at genscan@gnomic.stanford.edu), searches
            of the EST database at TIGR  (http://www.tigr.org/tdb/hcd/hcd.html)
            and searches against a peptide database. Repeats were identified
            using RepeatMasker (Smit, A. and Green,  P. unpublished,
            http://ftp.genome.washington.edu/rm/RepeatMasker.html).
FEATURES             Location/Qualifiers
     source          1. .202004
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16p12"
                     /clone="A-270G1"
     mRNA            join(46775. .46931,47279. .47384,48522. .48578,53429. .53518,
                     53625. .53754,54203. .54262,55202. .55449)
                     /gene="A-270G1.2"
     gene            46775. .55449
                     /gene="A-270G1.2"
     CDS             join(47289. .47384,48522. .48578,53429. .53518,53625. .53702)
                     /gene="A-270G1.2"
                     /codon_start=1
                     /product="Unknown gene product"
                     /protein_id="AAC27823.1"
                     /db_xref="GI:3355304"
                     /translation="MRALENDFFNSPPRKTVRFGGTVTEVLLKYKKGEINDFELLKNQ
                     LLDPDIKDDQIINWLLEFRSSVMYLTKDFEQLISIILRLPWLNRSQTVVEEYLAFIGN
                     LVSA"
     STS             54435. .54535
                     /gene="A-270G1.2"
                     /db_xref="dbSTS:G27879"
     mRNA            join(63313. .63405,65204. .65334,66207. .66391,68184. .68291,
                     68925. .69073)
                     /gene="A-270G1.1"
     gene            63313. .69073
                     /gene="A-270G1.1"
     CDS             join(63403. .63405,65204. .65334,66207. .66391,68184. .68291,
                     68925. .68958)
                     /gene="A-270G1.1"
                     /codon_start=1
                     /product="Unknown gene product"
                     /protein_id="AAC27822.1"
                     /db_xref="GI:3355303"
                     /translation="MGFAEAFLEHLWKNLQDPSNPAIIRQAAGNYIGSFLARAKFISL
                     ITVKPCLDLLVNWLHIYLNNQDSGTKAFCDVALHGPFYSACQAVFYIFVFRHKQLLSG
                     NLKEGLQYPQSLNFERIVMSQLNPLKICLPSVVNFFAAIINRMKTCGYGWW"
     STS             174745. .174876
                     /db_xref="dbSTS:G19949"
     STS             184872. .185077
                     /db_xref="dbSTS:G04856"
BASE COUNT    56977 a  45624 c  44447 g  54956 t
ORIGIN


  Query Match             8.5%;  Score 176;  DB 9;  Length 202004;
  Best Local Similarity   100.0%;  Pred. No. 3e-81;
  Matches  176;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1339 catgcctagatcttttggttaactggctgcacatatacctaataaccaggattcgggaa 1398
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66278 CATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCAGGATTCGGGAA 66277
```

```
Qy    1399 caaaqqcattctqcqatqttqctctccatqqaccatttactcaqcctqccaaqctgtqt 1458
           ||||||||||||||||||||||||||||||||||||.||||||||||||||||||||||
Db   66278 CAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGT 66337

Qy    1459 tctqcacctttqttttttaqacacaaqcaqctttqaqcqqaaacctqaaaqaaqqt 1514
           ||||||||||||||||||||||||||||||||||.|||||||||||||||||||
Db   66338 TCTACACCTTTGTTTTTAGACACAAGCAGCTTTGAGCGGAAACCTGAAAGAAGGT 66393


Search completed: July 30, 2002, 05:34:49
Job time: 8689 sec
```

OM of: US-09-932-678-2 to: EST:* out_format : pfs

Date: Jul 30, 2002 5:45 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09932678/runat_29072002_155123_17917/app_query.fasta_1.717
-DB=EST -QFMT=fastap -SUFFIX=std.rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT pfs -NOKM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=2000000000 -USER=US09932678_@CGN1_1_3018
-NCPU 6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT 30
-NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-932-678-2
Query length: 651
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1680.260000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
qb_est2:Y16973	+	1755.50	3477.78	1.8e-184	1800	! Y16973 RNY16973 Rat liver ESTs
qb_est2:BG502871	+	1207.50	2389.90	7.2e-124	788	! BG502871 602550430F1 NIH_MGC_61
qb_est2:BF203684	+	1125.50	2223.03	1.4e-114	996	! BF203684 601866331F1 NIH_MGC_17
qb_est2:BI561523	+	1105.00	2185.48	1.7e-112	733	! BI561523 603256222F1 NIH_MGC_97
qb_est2:BG428305	+	1100.50	2172.51	9.2e-112	1037	! BG428305 602498983F1 NIH_MGC_7
qb_est2:BE734354	+	1094.00	2163.08	3.1e-111	757	! BE734354 601565596F1 NIH_MGC_21
qb_est1:AV703279	+	1040.00	2056.33	2.7e-105	671	! AV703279 AV703279 ADB Homo sapi
qb_est2:BI761160	+	1000.50	1974.84	9.4e-101	828	! BI761160 603043657F1 NIH_MGC_11
qb_est2:BI464721	+	989.50	1954.16	1.3e-99	736	! BI464721 603202341F1 NIH_MGC_97
qb_est1:AW408066	+	966.00	1910.15	3.8e-97	563	! AW408066 UI-HF-BM0-adu-b-10-0-U
qb_est2:BI549394	+	966.00	1907.24	5.5e-97	727	! BI549394 603190049F1 NIH_MGC_95
qb_est2:BG167061	+	935.00	1841.65	2.5e-93	990	! BG167061 602344740F1 NIH_MGC_83
qb_est2:BG297178	+	928.50	1828.02	1.4e-92	1045	! BG297178 602395174F1 NIH_MGC_9
qb_est1:AW958173	+	902.00	1781.98	5.2e-90	564	! AW958173 EST370243 MAGE reseque
qb_est2:BG295610	+	898.50	1768.96	2.8e-89	956	! BG295610 602392992F1 NIH_MGC_94
qb_est2:BI601238	+	892.00	1757.49	1.2e-88	835	! BI601238 603245755F1 NIH_MGC_96
qb_est1:BB642011	+	891.00	1758.73	1.0e-88	628	! BB642011 BB642011 RIKEN full-le
qb_est1:AA213789	+	874.00	1726.52	6.4e-87	535	! AA213789 zr91a11.s1 NCI_CGAP_GC
qb_est2:BE870692	+	871.00	1716.78	2.2e-86	742	! BE870692 601448528F1 NIH_MGC_65
qb_est1:BB639379	+	860.00	1695.65	3.3e-85	686	! BB639379 BB639379 RIKEN full-le
qb_est1:AL597171	+	846.00	1670.13	8.8e-84	550	! AL597171 DKFZp313K0411_r1 313 (
qb_est1:AA530643	+	840.00	1658.18	4.1e-83	547	! AA530643 vj44a04.r1 Stratagene
qb_est2:BI913245	+	835.00	1643.74	2.6e-82	807	! BI913245 603178986F1 NIH_MGC_12
qb_est2:BG535886	+	826.50	1624.91	2.9e-81	946	! BG535886 602564040F1 NIH_MGC_77
qb_est2:BF221544	+	826.00	1631.24	1.3e-81	497	! BF221544 7o69d12.x1 NCI_CGAP_Pr
qb_est1:AW968124	+	823.00	1624.52	3.1e-81	529	! AW968124 EST380319 MAGE reseque
qb_est2:BG611364	+	821.00	1617.38	7.7e-81	697	! BG611364 602612725F1 NIH_MGC_60
qb_est2:BF110176	+	820.00	1619.20	6.1e-81	498	! BF110176 7n50d06.x1 NCI_CGAP_La
qb_est2:BG149493	+	812.00	1603.19	4.7e-80	498	! BG149493 nad29e03.x1 NCI_CGAP_L
qb_est1:AI799954	+	804.00	1587.05	3.7e-79	503	! AI799954 wc41d02.x1 NCI_CGAP_Pr
qb_est2:BE502966	+	799.00	1577.34	1.3e-78	490	! BE502966 hz81q07.x1 NCI_CGAP_La
qb_est2:BE560866	+	784.50	1542.57	1.1e-76	811	! BE560866 601346151F1 NIH_MGC_8
qb_est2:BE549643	+	781.00	1542.07	1.2e-76	458	! BE549643 7b40g08.x1 NCI_CGAP_Lu
qb_est2:BE614096	+	770.00	1516.37	3.2e-75	632	! BE614096 601503834F1 NIH_MGC_71
qb_est1:AI928274	+	767.00	1513.96	4.4e-75	461	! AI928274 wo67c08.x1 NCI_CGAP_Pr
qb_est1:AA811628	+	759.00	1498.47	3.2e-74	440	! AA811628 ob74d03.s2 NCI_CGAP_GC
qb_est2:BE271211	+	748.50	1470.54	1.2e-72	807	! BE271211 600943821F1 NIH_MGC_8
qb_est1:BE093077	+	741.00	1462.90	3.1e-72	422	! BE093077 RC5-BT0744-260400-031
qb_est2:BG863473	+	738.00	1446.42	2.6e-71	1059	! BG863473 602796946F1 NIH_CGAP_
qb_est1:BB665390	+	736.50	1448.13	2.0e-71	700	! BB665390 BB665390 RIKEN full-le
qb_est1:AA191111	+	732.00	1444.91	3.1e-71	421	! AA191111 zp88a03.r1 Stratage
qb_est2:BI225531	+	678.50	1332.17	5.9e-65	689	! BI225531 602948775F1 NIH_MGC
qb_est1:AA481295	+	673.00	1327.94	1.0e-64	380	! AA481295 zv62e05.r1 NCI_CGAP
qb_est2:BI018122	+	664.00	1309.16	1.1e-63	406	! BI018122 CMI-M10244-020801 7
qb_est2:BI914307	+	664.00	1303.63	2.3e-63	660	! BI914307 602880944F1 NIH_MGC

seq_name: gb_est2:Y16973

seq_documentation_block:
LOCUS Y16973 1800 bp mRNA linear EST 30-APR-1999
DEFINITION RNY16973 Rat liver ESTs (E.Olivier) Rattus norvegicus cDNA clone IRL573, mRNA sequence.
ACCESSION Y16973
VERSION Y16973.2 GI:4727035
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 1800)
AUTHORS Olivier,E., Soury,E., Kisler,J.L., Smih,F., Schneider,K., Lochner ,K., Jouzeau,J.Y., Fey,G. and Sallot,J.P.
TITLE A novel set of hepatic mRNAs preferentially expressed during an acute inflammation in rat represents mostly intracellular proteins
JOURNAL Genomics 57 (3), 352-364 (1999)
MEDLINE 99263497
COMMENT On Apr 2, 1998 this sequence version replaced gi:3006369.
Contact: E. Olivier
U78 INSERM
543 chemin de la Breteque, 76233 Bois-Guillaume, France
POLYA=No.
FEATURES Location/Qualifiers
source 1. .1800
/organism="Rattus norvegicus"
/strain "Sprague-Dawley"
/db_xref="taxon:10116"
/clone="IRL573"
/clone_lib="Rat liver ESTs (E.Olivier)"
/tissue_type="liver"
/dev_stage="adult"
/note="Organ: liver; Vector: lambda ZAPII; Library construction: Ripperger,J.A. et al. J. Biol. Chem. (1995) 270(50):29998-30006"
BASE COUNT 473 a 393 c 433 g 501 t
ORIGIN

alignment_scores:
Quality: 1755.50 Length: 625
Ratio: 3.294 Gaps: 24
Percent Similarity: 85.280 Percent Identity: 69.600

alignment_block:
US-09-932-678-2 x Y16973 ..

Align seq 1/1 to: Y16973 from: 1 to: 1800

```
      24 LeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAs 40
         |||::::|||||||||||||||:::|||||||||   ||||||||||:::||
      26 CTGAGTGCGTTGCGGACTGGGCTTTCAAACATGCTTGCATTAGAGAGTGA 75

      41 pPhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValT 57
         ||||       |||       :::::|||||!   :::       ||||||
      76 TTTC......TCAGTCTCCCAAGAAACTGTCGGTTCG...TGGACTGTGA 116

      57 hrGluValLeuLeuLysTyrLysLysGlyGluThrAsnAspPheGluLeu 73
         ||:::|||||||||:::       ||||||   ||||||||||:::|||||
     117 CAGACGTCTTACTGAGT...ACAAAAGGGTGAACAAATGACTTAGAGTTG 163

      74 LeuLysAsnGlnLeuLeuAspProAspIleLysAspAspGlnIleIleAs 90
         ,|||||||   |||  ||||||||||||:::||||||||||||||||||
     164 TTGAAGAACAG.CTGTCTGATCCTGATATAGAGGATGATCAGATCATTAA 212
```

```
 90 nTrpLeuLeuGluPheArgSerSerIleMetTyrLeuThrLysAspPheG 107
    ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||
213 CTGGCTGCTGGAATTTCGTTCCTCTGTCATGTACTTGACAAAAGACTTTG 262

107 InGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgSerGln 124
    ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
263 AGCAACTTATAAACATCATATTGACATTGCCCTGGTTGAATAGAACTCAC 312

124 ThrValValGlnGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGl 140
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
313 ACGGTGGTGGAGGAGTATTTGCCTTTTCTTGGTAATCTTGTGTCAGCACA 362

140 nThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValP 157
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
363 GACTGTCTTCCTTAGACCATGTCTCAGCATGATTGCTTCTCATTTTGTAC 412

157 roProArgValIleIleLysGluGlyAspValAspValSerAspSerAsp 174
    |||||||||||||||:::|||||||||   :::|||   |||:::|||:::
413 CTCCCCGACGTTATTGTCAAGGAAGGTGGCATTGATGT.TCAGAGTCTGAG 461

174 AspGluAspAspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLe 190
    ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||
462 GATGAAGATGACAATCTTCCTGCAAATTTTGACACATGTCACAAAGCCTT 511

190 nGlnIleIleAlaArgTyrValProSerThrProTrpPheLeuMetProI 207
    |||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
512 GCAAATAATAACAAGATATGTCCTTCGACACCATGGTTTCTAATGCCAA 561

207 IeLeuValGluLysPheProPheValArgLysSerGluArgThrLeuGlu 223
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
562 TACTGGTAGAAAAATTTCCATTTGTGAGGAAATCAGACAGAACACTGGAA 611

224 CysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuAr 240
    ||||||||||||||||||||||||||||||||:::|||||||||||||||||
612 TGCTATGTTCATAACCTATTAAGCATAAGTTTATATTTCCCAACGTTGAG 661

240 qHisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspValA 257
    |:::|||||||||||||||||||||||:::   |||||||||||||||||:
662 GTGTGAAATTCTCGAGCTTATTATTGAAAGTCA.CTCAAGTTAGATGTGA 710

257 snAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThrGlnThr 274
    :::::|||||||||   |||:::|||||||||||||||||||:::||||||
711 GTGTGTTGCGCCAGGATATTGACGATGCTGAAGAGACAGCAGCTCACACT 760

274 CysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGluAspGl 290
    ||||||||||||:::::|||||||||||:::   :::
761 TATGCTGGGACAGACGCCCACAACGACTGTTA...ATATGGATGAGATG 807

290 nGluThrGluHisGluThrLysAlaGlyProGluArgLeuAspGlnMetV 307
    |||:::   |||   ||||||      |||||||||   :::|||:
808 AGGAACTGACCCTGAGAAGAAAGCTGAGCAGGAGGCGCTGGC.CGGATGG 856

307 alHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeuSer 324
    ::||||||:::|||||||||||||||||||||||:::|||||||:::||||||
857 CTCATCCCACTGCAGAGCCCTGGACATCCTGCTGTCCTTGCTGCTCTCC 906

324 TyrMetLysAspValCysTyrVal.AspGlyLysValAspAsnGlyLysT 340
    |||:::   ||||||||||:::||| :::   |||:::|||||:::|||||
907 TACATTGACGATGTCTGCCATGTGCAGT...AAAATTGATAACAATAAAA 953

340 hrLysAspLeu.TyrArgAspLeuIleAsnIlePheAspLysLeuLeuLe 356
    |||||   ::: |||||||||||||||||:::|||:::          :::||
954 CAAAG...ATTATATCGTGATCTCATATCTATCTTG...ACAACTGTGTT 997

356 nProThrHisAlaSerCysHisValGluPhePheMetPheTyrLeuCysS 374
    |||||||||||||   |||||||||||||||||||||||||:::|||:::::
998 GCCCACACATGCCTC.TGCCATGTACAGTTCTTCATGTTTTTCCTTGCAG 1046

374 crPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLys 389
    ::   ::::::|||   :::::|||||||||||||||||      |||
1047 CT...CAAATTGGGGTTTGCAGAGCATTTTTGGAACATCTC...TGAAAA 1090
```

```
390 IenGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTy 406
    :::|||||   :::|||||||:::|||||   :::::::::   :::||
1091 GTGCAGGATCA.AATAACCCTGTCATCATCAGCAGCTGCTGC...AAAIA 1136

406 rIleGlySerPheLeuAlaArgAlaLys.PheIleProLeuIleThrVal 422
    |:::   ||||||||||||:::   :::       ||||||||||||||||
1137 CATG...AGCTTTTTGGCAGAACTAAATTAAT...CCTCTTATCACTGTG 1180

423 LysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAs 439
    ||||||||||||||||      ||||||||||||      |||
1181 AAGTCATGCCTGGATCTTTGGGT.AACTGGCTGCACTGTACCTTAATAGC 1229

439 nGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProP 456
    :::   :::         ::::::|||                ||||||||||
1230 CAGGTTCGGACAAGGCTTTGTGATGTGC.........ACTCATGGACCAT 1270

456 heTyrSerAlaCysGlnAlaVal.PheTyrThrPheValPheArgHisLy 472
    ||   |||   ||||||||||||| ::::::             :::|||
1271 TTATCAGC...TGCCAAGCTGTGTCTACACTGTGTTTAGACACAGCACTT 1317

472 sGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerL 489
    :::          ||||:::|||:::          ||||||||||||||
1318 GAGTGCACCTGAGGAAGTCTACGG..........TATCTCCAGAGTC 1355

489 snAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCys 505
    |:|||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||
1356 TAAATTTGAACGCATTGTGCTGAGCCAGCTGAACCCACTGAAGATCTGC 1405

506 LeuProSerValValAsnPheAlaAlaIleThrAsnLysTyrGlnLe 522
    |||:||   |||   :::::::::   :::|||              |||||
1406 CTGCCACGTGTGTATTCCTTGCTGCATACAATA.........ATACAGCT 1446

522 nValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProV 539
    |   :::   :::   |||:::|||||   ||||||||||||||||||||
1447 GTGTCTGCTACAC...ATCATGGAGAGGAC.AATCGCCAGATGCTCCTG 1492

539 alIleArgSer.ThrAlaGlyGlyAspSerValGlnIleCysThrAsnPr 555
    |||||   :::|||||||||||||||||||||||||   ||||||||||||
1493 TTATC...GAGTACAGCTGGTGGAGACTCTGTGCAGACCTGCACCAACCC 1539

555 oLeuAspThrPhePhePheAspProCysValLeuLysArgSerLysL 572
    |||||||||||||||||||||||||||||||:::|||||||||||||||||
1540 ATTGGACACTTTCTTTCCCTTTGACCCTTGTATGCTTAAGAGGTCAAAGA 1589

572 ysPheIleAspProIleTyrGlnValTrpGluAspMetSerAlaGlnGlu 588
    |||:|||||||||||:::|||||||:::|||||||||||||||   |||||||
1590 ACTTCATCGATCCTGTCTATCAGGCTTGGGAAGACATGAGTGATGAAGAG 1639

589 LeuGlnGluPheLysLysProMetLysLysAspIleValGluAspGluAs 605
    |:|||||||||||||||||   :::|||:::::|||||||||||||||||
1640 CTGCAGGAGCTTAGGAAGCCCACT.AGAAAGGAGCTAGTGGAGGATGAAGA 1688

605 pAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIleGlyI 622
    |:|||||||||||||||:::::::::   |||:::|||||||||   |||:
1689 CGATGACTTTTTGAAGGCGAGTGCC...CAGAGTGACACAGTGGCTGGCC 1735

622 leThrProSerSerPheAspThrHisPheArgSerProSerSerSerVal 638
    ::|||||:::|||||||||||||||||||||||||   |||||||||||||||
1736 TTACCCCGAACTCCTTTGACACCCACTTTCGAGT.CCTTCCAGTAGTGTG 1784

639 GlySerProProVal 643
    ||||||||||||||||
1785 GGCTCCCCTCCTGTG 1799

seq_name: gb_est2:BG502871

seq_documentation_block:
LOCUS       BG502871                 788 bp    mRNA    linear   EST 27-MAR-2001
DEFINITION  602550430F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4658071 5',
            mRNA sequence.
ACCESSION   BG502871
VERSION     BG502871.1  GI:13464388
KEYWORDS    EST.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 788)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1451  row: 1  column: 08
            High quality sequence stop: 699.
FEATURES             Location/Qualifiers
     source          1. .788
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4658071"
                     /clone_lib="NIH_MGC_61"
                     /tissue_type="embryonal carcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcgaggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT      204 a    178 c    176 g    228 t      2 others
ORIGIN


alignment_scores:
            Quality: 1207.50             Length:    265
              Ratio:  4.792               Gaps:      5
 Percent Similarity:  95.094   Percent Identity: 92.453

alignment_block:
 US-09-932-678-2 x BG502871    ..

 Align seq 1/1  to: BG502871  from: 1  to: 788

      302 ArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLeuMe 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        2 CGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCGCCTGGACATCCTGAT 51

      318 tSerLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyLysV 335
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       52 GTCTTIGGTTTTGTCCTACATGAAGGATGTCTGCTATGTAGATGGTAAGG 101

      335 alAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIlePhe 351
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      102 TTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATCTTT 151

      352 AspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMe 368
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      152 GACAAACTCCTGTTGCCCACCCATGCCTCCTGCCATGTACAGTTTTTCAT 201

      368 tPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluH 385
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      202 GTTTTACCTCTGTAGTTTCAAATTGGGATTCGCAGAGGCATTTTTGGAAC 251

      385 isLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGln 401
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      252 ATCTCTGGAAAAAATTGCAGGACCCAAGTAATCCTGCCATCATCAGGCAG 301

      402 AlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIlePr 418
          ||||||||||||||||||||||||||||    ||||||||||||||||||
      302 GCTGCTGGAAATTATATTGGAAGCTTTNTGGCAAGAGCTAAATTTATTCC 451

      418 oLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisI 435
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      352 TCTTATTACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACA 401

      435 leTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAla 451
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      402 TATACCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTTGCT 451

      452 LeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheVa 468
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      452 CTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGT 501

      468 lPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnT 485
          ||||||||||||||||||||||||||||||||||||||||    ||||
      502 TTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGTTNGCAGT 551

      485 yrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnPro 501
          |||||||||||||||||||||||||||||||||||||||||||||||||
      552 ATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAATCCC 601

      502 LeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAs 518
          ||||||||||||    ||||||||||||||||||||||||||||||||||
      602 CTGAAGATTTGCTG.CCCTCAGTGGTTAACTTTTTTGCTGCAATCACAAA 650

      518 n.LysTyrGlnLeuValPheCysTyrThrIleIle.GluArgAsnAsnAr 534
          | ||||||||||||||||||||||||||||||||| ||||||||||||||
      651 TTAAGTACCAGCTGCTCTTCTGTTACACCATCATTGGAGAGGAACAATCG 700

      534 gGln.MetLeuProVal.IleArgSerThrAlaGlyGlyAsp...SerVa 549
          |||| |||:::   ||| |||   ||||||||||||||:::   |||::
      701 GCAGAATGGTTGCAGTCCATTGGGAGTACCGCTGGAGGAGAACTCAGTGC 750

      549 lGlnIleCysThrAsnProLeuAspThrPhePhePro 561
          ::::         ||||||:::::|||::|||||
      751 CGAATCTGGGCAAAACCCGGTGGAAACCCTTTTTCCC 787

seq_name: gb_est2:BF203684

seq_documentation_block:
LOCUS       BF203684                996 bp    mRNA    linear   EST 06-NOV-2000
DEFINITION  601866331F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4099398 5',
            mRNA sequence.
ACCESSION   BF203684
VERSION     BF203684.1  GI:11097270
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 996)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM967  row: n  column: 07
            High quality sequence stop: 684.
FEATURES             Location/Qualifiers
     source          1. .996
```

```
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone="IMAGE:4099398"
                     /clone_lib "NIH_MGC_17"
                     /tissue_type "rhabdomyosarcoma"
                     /lab_host "DH10B (phage-resistant)"
                     /note "organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Rong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT       348 a    211 c    231 g    216 t
ORIGIN


alignment scores:
            Quality: 1125.50              Length:    241
              Ratio:  4.893                Gaps:        2
 Percent Similarity:  95.436   Percent Identity: 95.021

alignment block:
 US-09-932-678-2 x BE203684   ..

 Align seq 1/1  to: BE203684  from: 1  to: 996

     379 AlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAs 395
         ||||||||||||||||||||||||||||||||||||||  ||||||||||||||
       1 GCAGAGGCATTTTGGAACATCTCTGGAAAAAATG.CAGGACCCAAGTAA 51

     396 nProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuA 412
         ||||||||||||||||||||||||||||||||||||||||||||||||||||  |
      52 TCCTGCCATCATCAGCCAGGCTGCTGGAAATTATATTGGAAGCTTTTCGG 101

     412 IaArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeu 428
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     102 CAAGAGCTAAATTTATTCCTCTTATTACTGTAAAATCATGCCTAGATCTT 151

     429 LeuValAsnTrpLeuHisIleTyrLeuAsnAsnGluAspSerGlyThrLy 445
         :::|||||||||||||||||||||||||||||||||||||||||||||||||
     152 GTGGTTAACTGGCTGCACATATACCTTAATAACGAGGATTCGGGAACAAA 201

     445 sAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnA 462
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     202 GGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAG 251

     462 IaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsn 478
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     252 CTGGTTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGAGCGGAAAC 301

     479 LeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleVa 495
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     302 CTCAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGT 351

     496 lMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnP 512
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     352 GATGAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACT 401

     513 hePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrIle 528
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     402 TTTTTGCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGCTACACCATC 451

     529 IleGluArgAsnAsnArgGlnMetLeuProValIleArgSerThrAlaGl 545
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     452 ATTGAGAGCAACAATCGCCAGATGCTGCCAGTCATTAGGAGTACCGCTGC 501

     546 yGlyAspSerValGlnIleCysThrAsnProLeuAsp.ThrPhePhePro 561
         |||||||||||||||||  ||||||||||||||||| |||||||||||||
     502 AGGAGACTCAGTGCAGATCTG.A'AAACCCGCTGGACAACCTTCTTCCCC 550

     562 PheAspProCysValLeuLysArgSerLysLysPheIleAspProIleTy 578
```

```
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     551 CTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTTA 600

     579 rGlnValTrpGlnAspMetSerAlaGluGluLeuGlnGluPheLysLysP 595
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     601 TCAGCTGTGGGAACACATCAGTGCTGAAGAGCTACAGGAGTTCAAGAAAA 650

     596 roMetLysLysAspIleValGluAspGluAspAspAspPheLeuLysGly 611
         |||||||||||||||||||||||||||||||||||||  |||  |||
     651 CCATGAAAAAGGACATAGTGGAAGATGAAGATGATGCATTCTGAAAG... 697

     612 GlyValProGlnAsnAspThr 618
         |||  |||  ||||||
     698 GAGAATGGCCAGATGGATACG 718

seq name: gb_est2:BI561523

seq_documentation block:
LOCUS       BI561523                 733 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603256222F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5298581 5',
            mRNA sequence.
ACCESSION   BI561523
VERSION     BI561523.1  GI:15448837
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 733)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11755  row: I  column: 06
            High quality sequence stop: 717.
FEATURES             Location/Qualifiers
     source          1..733
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:5298581"
                     /clone_lib "NIH_MGC_97"
                     /lab_host "DH10B"
                     /note="organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI XhoI (ATCGAG
                     ); oligo-dT primed using primer 5' TTTTTTTTTTTTTTTVN 3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT       205 a    142 c    171 g    214 t      1 others
ORIGIN


alignment scores:
            Quality: 1105.00              Length:    234
              Ratio:  4.784                Gaps:        2
 Percent Similarity:  98.718   Percent Identity: 98.718

alignment block:
 US-09-932-678-2 x BI561523   ..

 Align seq 1/1  to: BI561523  from: 1  to: 733
```

```
      1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSe 17
        |||||||||||||||||||||||||||||||||||||||||||||||||
     23 ATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGATGCGGCCGCTTC 72

     17 rSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnM 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     73 GTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATA 122

     34 etArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrVal 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    123 TGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTT 172

     51 ArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGlyGl 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    173 CGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGA 222

     67 uThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleL 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    223 AACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAA 272

     84 ysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    273 AGGATGACCAGATCATCAACTGGTTGCTAGAATTCCGTTCTTCTATCATG 322

    101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuPr 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    323 TACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCC 372

    117 oTrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLeuG 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    373 TTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTG 422

    134 lyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMet 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    423 GTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATG 472

    151 IleAlaSerHisPheValProProArgValIleIleLysGluGlyAspVa 167
        |||||||||||||||||||  ||||||||||||||||||||||||||||
    473 ATTGCTTCCCATTTTGTGCCTTCCCGAGTGATCATTAAGGAAGGCGATGT 522

    167 lAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPheA 184
        ||||||||||||||||||||||||||||||||||||||||||  ||||
    523 AGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCTGCAAA.TTTG 571

    184 spThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    572 ACACATGTCACAGAGCCTTGCAAATAATAGCAAGATATGTACCATCGACA 621

    201 ProTrp.PheLeuMetProIleLeu.ValGluLysProProPheValArg 216
        |||||| ||||||||||||||||| ||||||||||||||||||||||||
    622 CCGTGGGTTTCTCATGCCAATACTNGGTGGAAAAATTTCCATTTGTTCG. 670

    217 LysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIle 232
        ||||||||||||||||||||||||||||||||||||||||||||||||
    671 AAATCAGAGAGAACACTGGAATGTTACGTTCATAACTTACTAAGGATT 718

seq_name: qb_est2:BG428305

seq_documentation_block:
LOCUS       BG428305                1037 bp    mRNA    linear   EST 14-MAR-2001
DEFINITION  602498983F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4612605 5',
            mRNA sequence.
ACCESSION   BG428305
VERSION     BG428305.1  GI:13334811
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1037)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1361  row: e  column: 22
            High quality sequence stop: 637.
FEATURES             Location/Qualifiers
     source          1..1037
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4612605"
                     /clone_lib="NIH_MGC_75"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcgagcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-d(T30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.65
                     kb (range 0.5-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT       292 a    218 c    263 g    264 t
ORIGIN


alignment_scores:
            Quality: 1100.50              Length:    300
              Ratio:  4.031                Gaps:      14
 Percent Similarity:  91.000   Percent Identity: 84.333

alignment_block:
 US-09-932-678-2 x BG428305   ..

 Align seq 1/1  to: BG428305  from: 1  to: 1037

      1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSe 17
        |||||||||||||||||||||||||||||||||||||||||||||||||
     22 ATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGATGCGGCCGCTTC 71

     17 rSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnM 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     72 GTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATA 121

     34 etArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrVal 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    122 TGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTT 171

     51 ArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGlyGl 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    172 CGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGA 221

     67 uThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleL 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    222 AACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAA 271

     84 ysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    272 AGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATG 321

    101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuPr 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    322 TACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCC 371

    117 oTrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLeuG 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
      372 TTGGTTGAATACAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTG 421

      134 lyAsnLeuValSerAlaGluThrValPheLeuArgProCysLeuSerMet 150
          |||||||||||||||||||||||||||||||||||||||||||||||||||
      422 GTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATG 471

      151 IleAlaSerHisPheValProProArg.ValIleIleLysGluGlyAspV 167
          |||||||||||||||||||||||||| ||||||||||||||||||||||
      472 ATTGCTTCCCATTTGTGCCTCCCGAAGTGATCATTAAGGAAGGCGATG 521

      167 alAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPhe 183
          |||||||||||||||||||||||||||||||||||||||||||||   |||
      522 TAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCTGCAAA.TTT 570

      184 AspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerTh 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||
      571 GACACATGTCACAGAGCCTTGCAAATAATAGCAAGATATGTACCATCGAC 620

      200 rPro.TrpPheLeuMetProIleLeu.Val.GluLysProProPhe.Val 215
          |||| ||||||||||||||||||||| ||| ::::::|||||||::: |||
      621 ACCGTTGGTTTCTCATGCCAATACTGGGTGGCAACAATTCCCATATGGTC 670

      216 ArgLysSer...GluArgThrLeuGlu.CysTyr.ValHis.AsnLeuLe 230
          |||||||::   ||||||||||||||| |||:: |||||| ||||||||
      671 AGAAAAACTCCAGACGAACACTGGAACTGTCACCGTTCATGAACTTACT 720

      230 u.ArgIleSerValTyrPheProThrLeu.ArgHis.GluIleLeuGluL 246
          | |||||||::   |||   |||:::||| |||||| :::||||||||||
      721 TCAGGATCACGGGATATGTCCCAAGCTTGGAGGCATGCAAATTCTGGAGC 770

      246 euIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGly 262
          ||||||||||||:::      :::      ||||||:::||||||   |||
      771 TTATTATTGAAAACTACTCAAGGTGGACTGTGAATGGCTCCCGGGGCGGG 820

      263 Ile...GluAspAla...GluGluThrAlaThrGlnThrCysGlyGlyTh 277
          :::   |||||||::      :::      :::||||||   :::||||||
      821 GTATTGAAGACTGCTGAGGAAAGCACGAAGTACAAACTGGTGCGGGAAC 870

      277 rAspSerThrGluGlyLeuPheAsnMet 286
          |:::      :::||||||||||:::::
      871 AAATCCCGGCAAGGGATTGTTCAGTTTG 898

seq name: qb_est2:BE734354

seq documentation block:
LOCUS       BE734354                757 bp    mRNA    linear   EST 15-SEP-2000
DEFINITION  601565596F1 NIH MGC_21 Homo sapiens cDNA clone IMAGE:3840623 5',
            mRNA sequence
ACCESSION   BE734354
VERSION     BE734354.1  GI:10148346
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 757)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM540  row: o  column: 24
            High quality sequence stop: 698.
FEATURES             Location/Qualifiers
     source          1..757
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:3840623"
                     /clone_lib "NIH MGC 21"
                     /tissue_type "choriocarcinoma"
                     /lab_host "DH10B (phage-resistant)"
                     /note="Organ: placenta; Vector: pOTB7; Site 1: XhoI;
                     Site 2: EcoRI; cDNA made by oligo dT priming;
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT       233 a    152 c    166 g    206 t
ORIGIN


alignment_scores:
            Quality: 1094.00              Length:    246
              Ratio:  4.636                Gaps:      4
 Percent Similarity:  95.935   Percent Identity: 90.244

alignment_block:
 US-09-932-678-2 x BE734354    ..

 Align seq 1/1  to: BE734354  from: 1  to: 757

      181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrVa 197
          |.|||||||||||||||||||||||||||||:::|||||||||||||||
        1 GCAAATTTGACACATGTCACAGAGCCTTGCAACTAATAGCAAGATATGT 52

      197 lProSerThrProTrp.PheLeuMetProIleLeuValGluLysPhePro 213
          |||||||||||||    |||||||||||||||||||||||||||||||||
        53 ATCATGGACACCGTGTATTTCTCATGCCAATACTGGTGGAAAAATTTCCA 102

      214 PheValArgLysSerGluArgThrLeuGlu.CysTyrValHisAsnLeuL 230
          |.|||||||||||||||||||||||||:: ||||||||||||||||||||
      103 TTTGTTCGAAAATCAGAGAACACTGGACATGTTACGTTCATAACTTAC 152

      231 euArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLeu 246
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      153 TAAGGATTAGTGTATATTTCCAACCTTGAGGCATGAAATTCTGGAGCTT 202

      247 IleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGly.I 263
          |||||||||||||||||||||||||||||||||||||||||||||||| |
      203 ATTATTGAAAAACTACTCAAGTTGGATGTGAATGCATCCCGGCAGGGCTA 252

      263 leGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSer 279
          |||||||||||||||||||||||||||||||||||||||||||||||||||
      253 TTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGGTGGGACAGATTCC 302

      280 ThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluTh 296
          |||||||||||||||||||||||||||||||||||||||||||||||||||
      303 ACGGAAGGATTGTTTAATATGGATGAAGATGAAGAAACTGAACATGAAAC 352

      296 rLys.AlaGlyProGluArgLeuAspGluMetValHisProValAlaGlu 312
          |::: ||||||||||||||||||||||||||||||||||||||||||||||
      353 AAACGGCTGGTCCTGAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAG 402

      313 ArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCy 329
          ||||||:|||||||||||||||||||||||||||||||||||||||||||
      403 CGCCTGGACATCCTGATGTCTTTGGTTTTGTCCTACATGAAGGATGTCTG 452

      329 styrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgA 346
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      453 CTATGTAGATGGTAAGGTTGATAACGGCAAAACAAAGGATCTATATCGCG 502

      346 snLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCys 362
          |||||||||||||||||||||||||||||   |||:::|||||||||||||
      503 ACCTGATAAACATCTTTGACAAACTCCTGTAGCCCAACCATGCCTCCTGC 552

      363 HisValGluPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheAl 379
          |:|||||:::::::|||:::|||||||||||||||||   ||||||||
      553 CATGTACAGTATTACATGTATTACCTCTGTAGTTTCAAATCGGGATTCGC 602
```

```
     379 aGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnP 396
         |||||||:::   ||||||||||||||||||||||||||||||||||||||
     603 AGAGGCATATTAGGAACATCTCTGGAAAAAATTGCAGGACCCAAGIAATC 652

     396 roAlaIleIleArqGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAla 412
         ||||||||||||||||||||||||||||||:::|||||   ::::::
     653 CTGCCATCATCAGGCAGGCTGCTGGAAATTATATGGGAAGCTT.ATGGGA 701

     413 ArqAlaLysPheIleProLeuIleThrVal 422
         ||||||   :::||||||||||:::   |||
     702 AGAGCCTAATTAATTCCTCTATTACTGGTC 731

seq_name: qb_est1:AV703279

seq_documentation_block:
LOCUS       AV703279                671 bp    mRNA    linear   EST 09-OCT-2000
DEFINITION  AV703279 ADB Homo sapiens cDNA clone ADBCCD12 5', mRNA sequence.
ACCESSION   AV703279
VERSION     AV703279.1  GI:10720608
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 671)
  AUTHORS   Peng,Y., Song,H., Huang,Q., Huang,C., Gu,Y., Yang,Y., Gao,G., Xiao
            ,H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z.,
            Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
            ,G., Hu,R., Chen,J., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA ADB clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzq@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .671
                     /organism="Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone="ADBCCD12"
                     /clone_lib="ADB"
                     /tissue_type="Adrenal gland"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      189 a    131 c    157 g    193 t      1 others
ORIGIN


alignment_scores:
            Quality: 1040.00             Length:    221
              Ratio:  4.860               Gaps:      1
 Percent Similarity:  96.833   Percent Identity: 93.213

alignment_block:
 US-09-932-678-2 x AV703279   ..

 Align seg 1/1  to: AV703279  from: 1  to: 671

     117 ProTrpLeuAsnArqSerGlnThrValValGluGluTyrLeuAlaPheLe 163
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       2 CCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCT 51

     133 uGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerM 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      52 TGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCA 101

     150 eIleAlaSerHisPheValProProArqValIleIleLysGluGlyAsp 166
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     102 TGATTGCTTCCCATTTGTGCCTCCCCGAGIGATCATTAAGGAAGGCGAT 151

     167 ValAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPh 183
         |||   ||||||||||||||||||||||||||||||||||||||||||||
     152 GTACATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCTGCAAATTT 201

     183 eAspThrCysHisArqAlaLeuGluIleIleAlaArgTyrValProSerT 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     202 TGACACATGTCACAGAGCCTTGCAAATAATAGCAAGATATGTACCATCGA 251

     200 hrProTrpPheLeuMetProIleLeuValGluLysPheProPheValArq 216
         |||||||||||||||||||||||||||||||||||||||||||||||||
     252 CACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTCGA 401

     217 LysSerGluArqThrLeuGluCysTyrValHisAsnLeuLeuArqIleSe 233
         |||||||||||||||||||||||||||||||||||||||||||||||||
     302 AAATCAGAGAGAACACTGGAATGTTACGTTCATAACTTACTAAGGATTAG 451

     233 rValTyrPheProThrLeuArqHisGluIleLeuGluLeuIleIleGluL 250
         |||||||||:::|||||||||||||||||||||||||||||||||||||
     352 TGTATATTTTCCAACCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAA 401

     250 ysLeuLeuLysLeuAspValAsnAlaSerArqGlnGlyIleGluAspAla 266
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     402 AACTACTCAAGTTGGATGTGAATGCATCCCGGCAGGGTATTGAAGATGCT 451

     267 GluGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLe 283
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     452 GAAGAAACAGCAACTCAAACTTGTGGTGGGACAGATTCCACGGAAGGATC 501

     283 uPheAsnMetAspGluAspGluGluThrGluHisGluThrLysAla.Gly 299
         |||||||||||||||||:::|||||||||||||:::|||:::|||
     502 GTTTAATATGGATGAAGATGACGAAACTGAACACTCAACACAGGCTGNTT 551

     300 ProGluArqLeuAspGlnMetValHisProValAlaGluArqLeuAspIl 316
         ||||||||||||||||||||||||||||||:::::|||||||||||||
     552 CCTGAACGGCTCGACCAGATGGTGCATCCTGTAGTCGACCGCCTGGACAT 601

     316 eLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspG 333
         ||||||||||:::   |||:::|||||||:::|||||||||   |:|||||
     602 CCTGATGTCTTTCGGTTTGGCCTACATGCAGGATGTCTGTAT.GTAGATG 650

     333 lyLysValAsp 336
         |||||   |||
     651 GTAAGGGTGAT 661

seq_name: qb_est2:BI761160

seq_documentation_block:
LOCUS       BI761160                828 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603043657F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5183895 5',
            mRNA sequence.
ACCESSION   BI761160
VERSION     BI761160.1  GI:15752738
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 828)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
```

```
            http://image.llnl.gov
            Plate: LLAM11459  row: a  column: 16
            High quality sequence stop: 822.
FEATURES              Location/Qualifiers
     source           1. .828
                      /organism "Homo sapiens"
                      /db_xref "taxon:9606"
                      /clone "IMAGE:5188895"
                      /clone_lib "NIH_MGC_116"
                      /lab_host "DH10B"
                      /note "Organ: pooled colon, kidney, stomach; Vector:
                      pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                      source anonymous pool of 3 colons, age 26 yo male, 49 yo
                      female, 71 yo male colon; 46 yo male kidney, and pool of 2
                      stomachs, 62 yo male and 70 yo female.  Library is
                      oligo dT primed and  directionally cloned (EcoRV site is
                      destroyed upon cloning).  Average insert size 1.4 kb,
                      insert size range 1 3 kb.  Library is normalized and
                      enriched for full-length clones and was constructed by C.
                      Gruber  (Invitrogen).  Research Genetics tracking code
                      024. Note: this is a NIH_MGC Library."
BASE COUNT       203 a    215 c    207 g    203 t
ORIGIN


alignment scores:
            Quality: 1000.50            Length:    225
              Ratio:  4.904              Gaps:        3
 Percent Similarity:  90.667   Percent Identity: 87.111

alignment block:
 US-09-932-678-2 x BI761160  ..

 Align seq 1/1  to: BI761160  from: 1  to: 828

     374 PheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLe 390
         ::::::  ||||||||||||||||||||||||||||||||||||||||||
     154 TATCGCACGGCATTCGCAGAGGCATTTTTGGAACATCTTTGGAAAAAATT 202

     390 nGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyr1 407
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     203 GCAGGATCCAAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATTATA 252

     407 leGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLys 423
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     253 TTGGAAGCTTTTTGGCAAGAGCTAAATTTATTCCTCTTATTACTGTAAAA 302

     424 SerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGl 440
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     303 CCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCA 352

     440 nAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheT 457
         |||||||||||||||||||||||||:::|||||||||||||||||||||
     353 GGATTCGGAACAAAGGCATTCTGCAATGTTGCTCTCCATGGACCATTTT 402

     457 yrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGln 473
         |||||||||||||||||||||||||||||||||||||||||||||||||
     403 ACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAG 452

     474 LeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAs 490
         ||||||||||||||||:|||||||||||||||||||||||||||||||||
     453 CTTTTGAGCGGAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAA 502

     490 nPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuP 507
         |||||||||||||||:|||||||||||||||||||||||||||||||||
     503 TTTGAGCGGATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCTGC 552

     507 roSerValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuVal 523
         ||:||||||||||||||||||||||||||||:::|||||||||||||||
     553 CCTCAGTGGTTAACTTTTTTGCTGCAATCACACATAACTACGAGTCGTC 602

     524 PheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProVal11 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||

     603 TTCTGCTACACCATCATCGAGAGGAACAATCGCCAGATGCTGCCAGTCAT 652

     540 eArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuA 557
         |:|||||||||:::|||||||||||||||||||   ||||||||||||||
     653 TACGACTACCACTGGAGGACACTCAGTGCAGACCTGCACAAACCCACTGG 702

     557 spThrPhePhe.ProPheAspProCysValLeuLysArgSerLysLysPh 574
         ||||||||    ||||||||||||||||||||||||||||||||:::
     703 ACACCTTCGTCCCCTTTGATCCCTGTGTGCTGAAGAGATCTTCAACTTG 752

     573 eIle.........AspProIleTyr............GluValTrpGluA 584
                      :::   ||||||           |||   |||||||
     753 CTGCTGAGCTAGGGAAGACATTTACTGGATCGGAACACAGAGTTGGGAGG 802

     583 spMetSerAlaGluGluLeuGln 590
         ||          :::   ||||||
     803 ACTCTGTCAGACAGATGTTACAA 825

seq_name: qb_est2:BI464721

seq_documentation_block:
LOCUS       BI464721                 736 bp    mRNA    linear   EST 21 AUG 2001
DEFINITION  603202341F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5268011 5',
            mRNA sequence.
ACCESSION   BI464721
VERSION     BI464721.1  GI:15255364
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 736)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11676  row: b  column: 12
            High quality sequence stop: 725.
FEATURES              Location/Qualifiers
     source           1. .736
                      /organism "Homo sapiens"
                      /db_xref "taxon:9606"
                      /clone: "IMAGE:5268011"
                      /clone_lib "NIH_MGC_97"
                      /lab_host "DH10B"
                      /note "Organ: testis; Vector: pBluescriptR (modified
                      pBluescript KS+); Site_1: BamHI; Site_2: SalI XhoI (qtcgaq
                      ); oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                      size-selected for  average insert size 2.2 kb and
                      normalized to ROT 5. This is a primary library enriched
                      for full-length clones and constructed using the
                      Cap-trapper method (Carninci, in preparation).  Library
                      constructed by M. Brownstein (NIMH/NHGRI, National
                      Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT       206 a    143 c    178 g    209 t
ORIGIN


alignment scores:
            Quality:  989.50            Length:    234
              Ratio:  4.477              Gaps:        4
 Percent Similarity:  94.444   Percent Identity: 89.744

alignment block:
```

```
US-09-932-678-2 x BI464721   ..

Align seq 1/1  to: BI464721  from: 1  to: 736

       1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      41 ATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGATGCGGCCGCTTC 90

      17 rSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnM 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      91 GTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATA 140

      34 etArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrVal 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     141 TGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTT 190

      51 ArgPheGlyGlyThrValThrGluValLeuLeuLysTyr..LysLysGly 66
         ||||||||||||||||||||||||||||||||||||||  :::::::|||
     191 CGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAACAACAGGGT 240

      67 GluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIl 83
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     241 GAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACAT 290

      83 eLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleM 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     291 AAAGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCA 340

     100 etTyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeu 116
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     341 TGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAACATTG 390

     117 ProTrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLe 133
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     391 CCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCT 440

     133 uGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerM 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     441 TGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCA 490

     150 etIleAlaSerHisPheValProProArg.ValIleIleLysGluGlyAs 166
         |||||:::|||||||||||||||||||| ||||||||||||||||||||
     491 TGATTGGTTCCCATTTTGTGCCTCCCCGAAGTGATCATTAAGGAAGGCGA 540

     166 pVal.AspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsn 182
         |||| |||||||||||||||||||||||||||||||||||||:::::::
     541 TGTAAGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCGTGCAAA 590

     183 Phe...AspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValPr 198
         |||     |||      ::::::|||:::|||||||||||||||||||||
     591 TTTGTGACAACAGTGTCACAGAGCCTTGAAATAATAGCAAGATATGTACC 640

     198 oSerThrProTrpPheLeuMetProIleLeuValGluLysPheProPheV 215
         |||||||   ||||||||||||   ||||||||||||||||||   :::
     641 ATCGACACG.TGGTTTCTCATGGCAATACTGGTGGAAAAATTCCATTTGT 689

     215 alArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeu 230
           :::|||||||||||||||||   :::|||||||||||||   ||||||
     690 TCGAAAAATCAGAGAGAACATGGAAATGTTACGTTCATTAATTACTA 736

seq_name: gb_est1:AW408066

seq_documentation_block:
LOCUS       AW408066                563 bp    mRNA    linear   EST 16-FEB-2000
DEFINITION  UI-HF-BM0-adu-b-10-0-UI.r1 NIH_MGC_38 Homo sapiens cDNA clone
            IMAGE:3062754 5', mRNA sequence.
ACCESSION   AW408066
VERSION     AW408066.1  GI:6927123
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 563)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Eco RI site shown at the beginning of the sequence.
             Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: M.B. Soares Lab
             cDNA Library Arrayed by: M.B. Soares Lab
             DNA Sequencing by: M.B. Soares Lab
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .563
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3062754"
                     /clone_lib="NIH_MGC_38"
                     /tissue_type="lymph"
                     /cell_type="germinal center B cells"
                     /cell_line="MGC85"
                     /lab_host="DH10B (LTI)"
                     /note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
                     Constructed from size fractionated cytoplasmic mRNA
                     (2.5-3.5kb). Directionally cloned. Cells provided by Louis
                     M. Staudt, Ph.D. Library preparation by Maria de Fatima
                     Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT      151 a    122 c    119 g    171 t
ORIGIN

alignment_scores:
            Quality:  966.00            Length:    185
              Ratio:  5.250               Gaps:      1
 Percent Similarity:  99.459   Percent Identity: 99.459

alignment_block:
 US-09-932-678-2 x AW408066   ..

 Align seq 1/1  to: AW408066  from: 1  to: 563

     330 TyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgAs 346
         |:|||||||||||||||||||||||||||||||||||||||||||||||||
      10 TATGTAGATGGTAAGGTTGATAACGGCAAAACAAAGGATCTATATCGCGA 59

     346 pLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSerCysH 363
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      60 CCTGATAAACATCTTTGACAAACTCCTGTTGCCCACCCATGCCTCCTGCC 109

     363 isValGlnPheMetPheTyrLeuCysSerPheLysLeuGlyPheAla 379
         |||||||||||||||||||||||||||||||||||||||||||||||||
     110 ATGTACAGTTTTTCATGTTTTACCTCTGTAGTTTCAAATTGGGATTCGCA 159

     380 GluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnPr 396
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     160 GAGGCATTTTTGGAACATCTCTGGAAAAAATTGCAGGACCCAAGTAATCC 209

     396 oAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaA 413
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     210 TGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAA 259

     413 rgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeu 429
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     260 GAGCTAAATTTATTCCTCTTATTACTGTAAAATCATGCCTAGATCTTTTG 309

     430 ValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAl 446
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     310 GTTAACTGGCTGCACATATACCTTAATAACCAGGATTCGGGAACAAAGGC 359

     446 aPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaV 463
```

```
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    360 ATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTG 409

    463 alPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeu 479
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    410 TGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTG 459

    480 LysGluGlyLeuGln.TyrLeuGlnSerLeuAsnPheGluArgIleValM 496
        |||||||||||||| |||||||||||||||||||||||||||||||||||
    460 AAAGAAGGTTTGCAGGTACCTTCAGAGTCTGAATTTTGAGCGGATAGTGA 509

    496 etSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAsnPhe 512
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    510 TGAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTT 559

    513 Phe 513
        |||
    560 TTT 562
```

seq_name: gb_est2:BI549394

```
seq_documentation_block:
LOCUS       BI549394                727 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603190049F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5261100 5',
            mRNA sequence.
ACCESSION   BI549394
VERSION     BI549394.1  GI:15436706
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 727)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11658  row: b  column: 13
            High quality sequence stop: 715.
FEATURES             Location/Qualifiers
     source          1..727
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone="IMAGE:5261100"
                     /clone_lib "NIH_MGC_95"
                     /tissue_type "hippocampus"
                     /lab_host "DH10B"
                     /note "Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgac
                     ); Oligo dT primed using primer 5'-TTTTTTTTTTTTTTTTTVN-3',
                     size selected for  average insert size 2.5 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for  full length clones and constructed using the
                     Cap trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health).  Note: this is a NIH_MGC Library."
BASE COUNT      182 a    191 c    171 g    183 t
ORIGIN

alignment_scores:
           Quality:  966.00             Length:     197
             Ratio:  5.031                Gaps:       1
 Percent Similarity:  97.462   Percent Identity: 93.401

alignment_block:
 US-09-932-678-2 x BI549394    ..

 Align seq 1/1  to: BI549394  from: 1  to: 727

    374 PheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLe 390
        ::::::    ||||||||||||||||||||||||||||||||||||||||
     98 TATCGCACGGCATTCGCAGAGGCATTTTTGGAACATCTTTGGAAAAAATT 147

    390 uGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrI 407
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    148 GCAGGATCCAAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATTATA 197

    407 leGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLys 423
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    198 TTGGAAGCTTTTTGGCAAGAGCTAAATTTATTCCTCTTATTACTGTAAAA 247

    424 SerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGl 440
           |||||||||||||||||||||||||||||||||||||||||||||||
    248 CCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCA 297

    440 nAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheT 457
        |||||||||||||||||||||||||:::||||||||||||||||||||||
    298 GGATTCGGGAACAAAGGCATTCTGCAATGTTGCTCTCCATGGACCATTTT 347

    457 yrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGln 473
        ||||||||||||||||||||||||||||:::|||||||||||||||||||
    348 ACTCAGCCTGCCAAGCTGTGTTCTACACCTTAGTTTTTAGACACAAGCAG 397

    474 LeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAs 490
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    398 CTTTTGAGCGGAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAA 447

    490 nPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuP 507
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    448 TTTTGAGCGGATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCTGC 497

    507 roSerValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuVal 523
        ||||||||||||||||:::|||||||||||||||||||||||||||||||
    498 CCTCAGTGGTTAACTTTTTAGCTGCAATCACAAATAAGTACCAGCTCGTC 547

    524 PheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIl 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    548 TTCTGCTACACCATCATCGAGAGGAACAATCGCCAGATGCTGCCAGTCAT 597

    540 eArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuA 557
        |||||||||:::||||||||||||||||||||    |||||:::||||||
    598 TAGGAGTACCACTGGAGGAGACTCAGTGCAGACCTGCACAAAACCACTGG 647

    557 sp.ThrPhePheProPheAspProCysValLeuLysArg 569
        || |||:::    |||||||||||||||||||||||||
    648 ACAACTTACGTCCCCTTTGATCCCTGTGTGCTGAAGAGA 686
```

seq_name: gb_est2:BG167061

```
seq_documentation_block:
LOCUS       BG167061                990 bp    mRNA    linear   EST 06-FEB-2001
DEFINITION  602344740F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4455009 5',
            mRNA sequence.
ACCESSION   BG167061
VERSION     BG167061.1  GI:12673764
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 990)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
           Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM10247  row: o  column: 10
           High quality sequence stop: 665.
FEATURES             Location/Qualifiers
     source          1. .990
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4455009"
                     /clone_lib="NIH_MGC_89"
                     /tissue_type="hypernephroma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT       304 a    275 c    219 g    192 t
ORIGIN


alignment_scores:
            Quality:  935.00             Length:    187
              Ratio:  5.109                Gaps:      1
 Percent Similarity:  97.861   Percent Identity: 97.326

alignment_block:
 US-09-932-678-2 x BG167061   ..

 Align seq 1/1  to: BG167061  from: 1  to: 990

     466 ThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGl 482
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       2 ACCTTTGTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGG 51

     482 yLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnL 499
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      52 TTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGC 101

     499 eu.AsnProLeuLysIleCysLeuProSerValValAsnPhePheAlaAl 515
         || :::||||||||||||||||||||||||||||||||||||||||||||
     102 TAAAGTCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGC 151

     515 alleThrAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgA 532
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     152 AATCACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGA 201

     532 snAsnArgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSer 548
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     202 ACAATCGCCAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCA 251

     549 ValGlnIleCysThrAsnProLeuAspThrPhePheProPheAspProCy 565
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     252 GTGCAGATCTGCACAAACCCGCTGGACACCTTCTTCCCCTTTGATCCCTG 301

     565 sValLeuLysArgSerLysLysPheIleAspProIleTyrGlnValTrpG 582
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     302 TGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTTATCAGGTGTGGG 351

     582 luAspMetSerAlaGluGluLeuGlnGluPheLysLysProMetLysLys 598
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     352 AAGACATGAGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAAG 401

     599 AspIleValGluAspGluAspAspAspPheLeuLysGlyGluValProGl 615
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     402 GACATAGTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCA 451

     615 nAsnAspThrValIleGlyIleThrProSerSerPheAspThrHisPheA 632
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     452 GAATGATACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTTCC 501

     632 rgSerProSerSerSerValGlySerProProValLeuTyrMetGlnPro 648
         ||||||||||||||||||||||||||||||   |||   |||||||||
     502 GAAGTCCTTCAAGTAGTGTGGGCTCCCCAACCGTGTAGTACATGCAAGCC 551

     649 SerProLeu 651
         |||||||||
     552 AGTCCCCTC 560

seq_name: qb_est2:BG297178

seq_documentation_block:
LOCUS       BG297178                1045 bp    mRNA    linear   EST 21 FEB 2001
DEFINITION  602395174F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4506594 5',
            mRNA sequence.
ACCESSION   BG297178
VERSION     BG297178.1  GI:13060570
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1045)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10382  row: d  column: 19
            High quality sequence stop: 687.
FEATURES             Location/Qualifiers
     source          1. .1045
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4506594"
                     /clone_lib "NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host "DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT       303 a    225 c    273 g    243 t      1 others
ORIGIN


alignment_scores:
            Quality:  928.50             Length:    336
              Ratio:  3.613                Gaps:     10
 Percent Similarity:  76.488   Percent Identity: 64.881

alignment_block:
 US-09-932-678-2 x BG297178   ..

 Align seq 1/1  to: BG297178  from: 1  to: 1045

       1 MetAlaAlaPro.LeuLeuHisThrArgLeuProGlyAspAlaAlaAlaS 17
         ||||||||:|||| ||||||||||||||||||   |||||::::::|||:
      26 AtGGCGGCCCCGGCTGCTTCACACGCGCTTGTCGGGCGATGTGACGGCAG 76

      17 erSerSerAlaValLysLysLeuGlyAlaSerArgThr.GlyIleSerAs 33
         :::::|||||:::      |||::::|||||||||| |||::|||::
      77 CGGCCTCTGCCACG......CTGAGTGCGTCGCGGACGTGGGCTTTCCGA 122
```

```
 33 nMetArqAlaLeuGluAsnAspPhePheAsnSerProProArqLysThrV 50
    :|||   |||||||||:::|||||||||||||||||||||||:::|||||||
123 FATGCTCGCGTTAGAGAGCGATTTCTTCAATTCTCCTCCGAAAAAAACTG 172

 50 nlArqPheGlyGlyThrValThrGluValLeuLeu.LysTyrLysLysGl 66
    |||||||||||||||||||||||||||||||||| ||||||||||||||
173 TTCGGTTTGGCGGAACTGTGACAGAAGTCTTACTGGAAATACAAAAAGGG 222

 66 yGlnThrAsnAspPheGluLeuLeuLysAsnGluLeuLeuAspProAspI 83
    |||||||||||||:::|||||||||||||||||||||   ||||||||||
223 TGAAACAAATGACTTAGAGTTGTTGAAGAACCAGCTGTCTGATCCTGATA 272

 83 IeLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArqSerSerIle 99
    |||||||||||||||||||||||||||||||||||||||||||||||:::
273 TAAAGGATCACCAGATCATTAACTGGCTACTAGAATTTCGTTCCTCTGTC 322

100 MetTyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArqIe 116
    |||||||||||||||||||||||||||||||||:::|||||||||||||
323 ATGTACTTGACAAAAGACTTTGAGCAACTTATAAACATCATATTGAGATT 372

116 nProTrpLeuAsnArqSerGlnThrValValGluGluTyrLeuAlaPheL 133
    |||||||||||||||||||    ||||||||||||||||||||||||||
373 GCCCTGGTTGAATAGAAGTCAGAGAGTGGTGGAGGAGTATTTGGCTTTTC 422

133 nnGlyAsnLeuValSerAlaGluThrValPheLeuArqProCysLeuSer 149
    |||||||||||||||||||||||||||||||||||||||||||||||||
423 TTGGTAATCTTGTGTCTGCACAGACTGTCTTCCTTACGACCATGTCTCAGC 472

150 MetIleAlaSerHisPheValProProArqValIleIleLysGlu.GlyA 166
    |||||||||||||:::|||||||||||||||||:::|||||| |||
473 ATGATTGCGTCTCATTTGGTACCTCCCCGAGTAATTGTCAAGGAACGGTG 522

166 spValAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsn 182
      :::|||||||||||||||||||||||||||||||||||||||    |||
523 GCATAGATGTTTCAGATTCTGATGACGAAGATGACAATCTTCCTGC.AAT 571

183 PheAspThrCysHisArqAlaLeuGlnIleIleAlaArqTyrValProSe 199
    |||||||||||||||||||||||||||||||:::|||||||||||||||
572 TTTGACACATGTCACAGAGCCTTGCAAATAATAACAAGATATGTCCCATC 621

199 rThrProTrpPheLeuMetProIleLeuValGluLysPheProPheValA 216
    |||||||||||||||||||||||||||||||:::::::|||    ||||
622 GACACCATGGTTTCTAATGCCAATACTGGTGGAGGAAATTCCGTT.GTGA 670

216 rqLysSerGluArqThrLeuGluCysTyrValHisAsnLeuLeu..... 230
    ||||||||||||||||||:::   :::|||||||||
671 GGAAGTCCGAGAGAACATTGACGGTCATGTTCATAATTATACGCATAGT 720

231 .....ArqIleSerVa 234
                  |||:::|||||
721 TTAATTTCAATTTGAGGGTGACATCTGGGGTGGTATTGAAGAGTATCAGT 770

234 TlyrPhePro.....ThrL 239
    |    ||||
771 TACACTTGACGTGTTTCGGGGACGACATGCAACGCGAAACAACCAGGCACAC 820

239 nnArqHisGluIleLeuGluLeuIleIle..... 248
    |||||:::      :::     :::
821 TCCGGCGCACGCACCACCAAGGACGGTAAAATGGCGCAGAACATAGGAAT 870

249 .....GluLysLeuLeuLysLeuAspValAsnAlaSerArqGlnGlyIl 263
         ||||||      :::      :::::|||:::||||||
871 GACACTGAAAAACAGTCACGGGCGCATACCAAGGCAACGAGACAA..... 915

263 nGluAspAlaGlnGluThrAlaThrGlnThrCysGlyGlyThrAspSerT 280
       |||   ::::::     ||||||     ::::::   |||:::
916 ....GACAGAAATCGCCATCGAACACAAGAAGACGCGGCGCAAAGACACAA 961

280 hrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThr 296
    |||:::          :::      :::||||||||||:::|||
962 AAGAAGCG.............ACAAAGAGAGGACACACAGAACATCACACA 999
```

```
seq_name: gb_est1:AW958173

seq_documentation_block:
LOCUS       AW958173                 564 bp    mRNA    linear   EST 01-JUN-2000
DEFINITION  EST370243 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW958173
VERSION     AW958173.1  GI:8147856
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 717
            Seq primer: Reverse.
FEATURES             Location/Qualifiers
     source          1..564
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGE"
                     /note="Vector: pBluescriptSKII-"
BASE COUNT       157 a    108 c    140 g    168 t      1 others
ORIGIN


alignment_scores:
            Quality:  902.00              Length:    188
              Ratio:  4.902                 Gaps:      1
 Percent Similarity:  97.872   Percent Identity: 97.872

alignment_block:
 US-09-932-678-2 x AW958173    .

 Align seq 1/1  to: AW958173  from: 1  to: 564

       6 LeuHisThrArqLeuProGlyAspAlaAlaAlaSerSerSerAlaValLy 22
         |:|||||||||:|||||||||||||||||||||||||||||||||||||
       3 CTTCACACGCGTTTGCCGGGAGATGCGGCCGCTTCGTCCTCTGCAGTTAA 52

      22 nLysLeuGlyAlaSerArqThrGlyIleSerAsnMetArqAlaLeuGluA 39
         |||||||||||||||||||||||||||||||||||||||||||||||||
      53 GAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATATGCGTGCATTAGAGA 102

      38 snAspPhePheAsnSerProProArqLysThrValArqPheGlyGlyThr 55
         |||||||:|||||||||||||||||||||||||||||||||||||||||
     103 ATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTTCGGTTTGGTGGAACT 152

      56 ValThrGluValLeuLeuLysTyrLysLysGlyGluThrAsnAspPheGl 72
         |||||||||||||||||||||||||||||||||||||||||||||||||
     153 GTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGA 202

      72 nLeuLeuLysAsnGlnLeuLeuAspProAspIleLysAspAspGluIleI 89
         |||||||||||:|||||||||||||||||||||||||||||||||||||
     203 GTTGTTGAAGAACCAGCTGTTAGATCCAGACATAAAGGATGACGAGATCA 252

      89 leAsnTrpLeuLeuGluPheArqSerSerIleMetTyrLeuThrLysAsp 105
         |||||||||||:|||||||||||||||||||||||||||||||||||||
     253 TCAACTGGCTCCTAGAATTCCGTTCTTCTATCATGTACTTGACAAAAGAC 302

     106 PheGluGlnLeuIleSerIleIleLeuArqLeuProTrpLeuAsnArqSe 122
         |:|||||||||||||||||||||||||||||||||||||||||||||||
```

```
     303 TTTGAGCAACTTATCAGTATTATATTAAGATTGCCTTGGTTGAATAGAAG 352

     122 rGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerA 139
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     353 TCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTGGTAATCTTGTATCAG 402

     139 laGlnThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPhe 155
         |||||||||||||||||||||||||||||||||||||||||||    ||||||
     403 CACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTTCCATTTT 452

     156 ValProProArgValIleIle.LysGluGlyAspValAspValSerAspS 172
         |||||||||||||||||||| ||||||||||||||||||||||||||||||
     453 GTGCCTCCCCGAGTGATCATTAAAGGAAGGCGATGTAGATGTTTCAGATT 502

     172 erAspAspGluAspAspAsnLeuProAlaAsnPheAspThrCysHisArg 188
         ||||||||||||||||||||||    |||    ||||||||||||||||||
     503 CTGATGATGAACATGATAATCTTCT.GCANATTTTGACACATGTCACAGA 551

     189 AlaLeuGlnIle 192
         ||||||||||||
     552 GCCCTGCAAATA 563

seq_name: qb_est2:BG295610

seq_documentation_block:
LOCUS       BG295610                 956 bp    mRNA    linear   EST 21-FEB-2001
DEFINITION  602392992F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4504751 5',
            mRNA sequence.
ACCESSION   BG295610
VERSION     BG295610.1  GI:13057417
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 956)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10377  row: q  column: 24
            High quality sequence stop: 660.
FEATURES             Location/Qualifiers
     source          1. .956
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4504751"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      266 a    204 c    239 g    247 t
ORIGIN


alignment_scores:
            Quality:  898.50             Length:    277
              Ratio:  3.682                Gaps:      8
 Percent Similarity:  88.087   Percent Identity: 71.841

alignment_block:
```

```
US-09-932-678-2 x BG295610   ..

Align seq 1/1  to: BG295610  from: 1  to: 956

    147 CysLeuSerMetIleAlaSerHisPheValProProArgValIleIleLy 163
        |||||||||||||||||||||||    ||||||||||||||    :::||
      4 TGTCTCAGCATGATTGCGTCTCATTT.GTACCTCCCCGAGTAAT.GTCAA 51

    163 sGluGlyAspValAspValSer.AspSerAspAspGluAspAspAsnLeu 179
        |||||||    :::|||||||| ||||||||||||||||||||||||||||
     52 GGAAGGTGGCATAGATGTTCAGGATTCTGATGACGAAGATGACAATCTA 101

    180 ProAlaAsnPhe.AspThrCysHisArgAlaLeuGlnIleIleAlaArgT 196
           :::|||||| ||||||||||||||||||||||||||||||:::|||||
    102 TCCTGCAATTTTGGACACATGTCACAGAGCCTTGCAAATAATAACAAGAT 151

    196 yrValProSerThrProTrpPheLeuMetProIleLeuValGluLysPhe 212
        |:|||||||||||||||||||||||||||||||||||||||||||||||||
    152 ATGTCCCATCGACACCATGGTTTCTAATGCCAATACTGGTGGAGAAATTC 201

    213 ProPheValArgLysSerGluArgThrLeuGluCysTyrValHisAsnIe 229
        |||   |||||||||||||||||||||||||||||||||||||||||||||
    202 CCGTT.GTGAGGAAGTCCGAGAGAACATTGGAATGTTATGTTCATAACTT 250

    229 uLeuArgIleSerValTyrProThrLeuArgHisGluIleLeuGluL 246
        |||||||||||||:::|||||||||||||||||||:::|||||||||||||
    251 ATTAAGGATAAGTTTATATTTCCCAACTTTGAGGCGTGAAATTCTGGAGC 300

    246 euIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGly 262
        |:::|||||||||||||||||||||||||||||:::::|||||||||||
    301 TTGTTATTGAAAAGCTACTCAAGTTAGATGTGAGTGTATCGCGGCAGGAT 350

    263 IleGluAspAlaGluGluThrAlaThrGluThrCysGlyGlyThrAspSe 279
        |:|||||||||||||||||    |||:::|||||||||||||||||||::
    351 ATTGAAGATGCTGAAGAGAAAGCAGCTCAGACTTGTGGTGGGACAGATAC 400

    279 rThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluT 296
        :|||||||||||||||||||||||||||||||||||:::|||:::   |||
    401 CACAGAAGGACTGTTTAATATGGATGAAGATGAGGACACTGACCCTGAGA 450

    296 hrLysAlaGlyProGluArgLeuAspGlnMetValHisProValAlaGlu 312
        ||||||     |||:::  :::|||||||:::|||||||:::|||||||
    451 ACAAAGCTGACCAGGAGCAGCCTAACCAGATGGCTCATCCCACCGCAGAG 500

    313 ArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCy 329
        |:|||||||:::|||:::   |||:::|||||||||:::::|||||||||
    501 CGCCTGGACGTCCTGCTGTGCTTGCTGCTGTCCTACATTGAGGATGTCTG 550

    329 sTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyrArgA 346
        |  |||   |||:'|:::|||||||:::|||||||||||||||||||||||
    551 CGGTGTGCACGGTAAAATTGATAACAATAAAACAAAGGATTTATACCGTG 600

    346 spLeuIleAsnIlePhe.AspLysLeu...LeuLeuProThrHisAlaSe 361
        ||||||||:::|||:: ||||||          ||||||||||||||||
    601 ATCTGATATCCATCTTGTGACAAAGCTTCTTGTTTGCCCACACATGCCTC 650

    361 rCysHisValGlnPhePhe...MetPheTyrLeuCysSerPheLysLeuG 377
        |||||||||      :::      |||:::|||   :::   :::::||
    651 CTGCCATGTACCGGTTCTTTCATGTTTTTTCCTCCGGCAGCCTTCAGTGG 700

    377 lyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspPro 393
        ||:::|||||||||:::||||||||||||||   ::::::   ||||||
    701 GCATCGCAGAAGCCATTTTGGAACATCTCTGGGCACAAGTGCCGGATCCA 750

    393 SerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPh 410
           :::      ||||||||||:::         :::|||:::|||||||
    751 AIAACC...CGCATAATCAGGAAG...CGGTGGAAATATTTGGGAAGCTT 794

    411 eLeuAlaArgAlaLysPheIlePro 418
        |   :::||||||::::::||||||
    795 T...GGCAGAGCTAGAATAATTCCC 816
```

OM of: US-09-932-678-2 to: Issued_Patents_NA:* out_format : pfs

Date: Jul 30, 2002 6:25 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09932678/runat_29072002_155124_17951/app_query.fasta_1.717
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=std.rni
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pis -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09932678_4CGN1_1_54 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-932-678-2
Query length: 651
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 53.260000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-329-681A-3	+	103.50	180.89	0.0218	2115	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-228-246-1	+	102.00	155.73	0.5499	10968	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-680-327-1	+	101.00	162.71	0.2246	5475	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-228-246-3	+	101.00	162.71	0.2246	5475	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-09-080-897-3	+	100.50	164.56	0.1772	4378	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-323-735-3	+	100.50	164.56	0.1772	4378	!
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-899-595-2	+	100.50	164.50	0.1786	4399	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-973-462-2	+	99.50	159.65	0.3327	5361	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-973-462-1	+	99.50	157.82	0.4204	6152	!
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-776-265-4	+	98.50	168.06	0.1131	2402	!
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-776-265-1	+	98.50	159.55	0.3369	4565	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-417-330A-13	+	97.00	169.79	0.0906	1638	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-680-327-2	+	97.00	144.59	2.29	10968	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-082-649B-80	+	96.50	171.14	0.0762	1360	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-082-649B-81	+	96.50	171.14	0.0762	1360	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-944-916-12	+	96.50	158.93	0.3648	3418	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-082-649B-83	-	96.50	156.89	0.4740	3987	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-082-649B-84	-	96.50	156.89	0.4740	3987	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-316-950-14	+	96.50	156.87	0.4752	3993	!
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-12642-14	+	96.50	156.87	0.4752	3993	
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-947-823-2	+	96.50	156.86	0.4760	3997	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-564-313-2	+	96.50	156.65	0.4886	4059	!
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US94-06069-2	+	96.50	156.65	0.4886	4059	
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-298-367B-3	-	96.50	156.63	0.4900	4066	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-977-630-20	+	96.50	156.45	0.5011	4120	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-977-630-21	-	96.50	156.45	0.5011	4120	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-316-950-11	+	96.50	156.45	0.5011	4120	!
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-12642-11	+	96.50	156.45	0.5011	4120	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-316-950-15	+	96.50	155.90	0.5381	4296	!
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-12642-15	+	96.50	155.90	0.5381	4296	
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-760-615-7	-	96.50	155.81	0.5445	4326	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-316-950-16	+	96.50	155.73	0.5501	4352	!
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-12642-16	+	96.50	155.73	0.5501	4352	
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-342-353-2	+	96.50	155.50	0.5663	4427	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-316-950-13	+	96.50	155.27	0.5833	4505	!
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-12642-13	+	96.50	155.27	0.5833	4505	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-345-913-1	+	96.50	154.08	0.6795	4928	!
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-818-562-1	+	96.50	154.08	0.6795	4928	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-564-313-1	-	96.50	153.98	0.6882	4965	!
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US94-06069-1	-	96.50	153.98	0.6882	4965	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-663-998-3	-	96.50	152.21	0.8039	5676	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-664-998-4	-	96.50	152.19	0.8055	5682	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-850-880-3	+	96.50	151.83	0.9063	5838	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-944-916-3	+	96.50	151.83	0.9063	5838	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-814-877-3	+	96.50	151.83	0.9063	5838	

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-329-681A-3

seq_documentation_block:
; Sequence 3, Application US/08329681A
; Patent No. 5650294
; GENERAL INFORMATION:
; APPLICANT: Kurth, Roland
; APPLICANT: Philippsen, Peter
; APPLICANT: Steiner, Sabine
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: No. 5650294el Promoter Region
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinkauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.4 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/329,681A
; FILING DATE: 26-OCT-1994
; CLASSIFICATION: 435
; CLASSIFICATION: C12P 25/00
; CLASSIFICATION: C12N 1/14
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2115 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-329-681A-3

alignment_scores:
Quality: 103.50 Length: 185
Ratio: 1.067 Gaps: 10
Percent Similarity: 52.432 Percent Identity: 27.027

alignment_block:
US-09-932-678-2 x US-08-329-681A-3 ..

Align Seq 1/1 to: US-08-329-681A-3 from: 1 to: 2115

```
      7 LysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAs 49
        |||:::|||   :::::::||   :::|||:::   |||      :::||
    880 AAAATGGGTAAGGAAAAGACTCACGTTTCGAGGCCGCGATTAAATTCCAA 929

     39 n.........AspPhePheAsnSerPro...ProArgLysThrVal...A 51
        |         |||::::::::::         |||   :::|||   :
    930 CATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGC 979

     51 rgPheGlyGlyThrVal.........................ThrGlu 58
        ::   |||:::|||:::                          |||
    980 AATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAG 1029

     59 ValLeuLeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLy 75
        ::::::|||||||::   ||||||:::::   :::
   1030 TTGTTTCTGAAACATGCCAAAGGTAGCGTTGCCAAT.............. 1065

     75 sAsnGlnLeuLeuAspProAspIleLysAspAspGlnIle...IleAsnI 91
```

```
                          |||:::    |||:::::::::    :::||||
1066 ....................GATGTTACAGATGAGATGGTCAGACTAAACT 1096

  91 rpLeuLeuGluPheArg......SerSerIleMetTyrLeuThrLysAsp 105
     |||||    ||||||               :::|||    ::::::   :::
1097 GGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACT 1146

 106 PheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgSe 122
         ::::::          :::          :::|||        :::
1147 CCTGATGATGCATGGTTACTCACCACTGCCGATCCCCGGGAAAACAGCATT 1196

 122 rGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerA 139
      |||    |||:::||||||||                |||:::|||:::|
1197 CCAG...GTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATG 1243

 139 laGlnThrValPheLeuArgProCysLeuSerMetIleAlaSerHis... 154
     ||     :::||||||||||||              |||:::   :::   :::
1244 CGCTGGCAGTGTTCCTGCGCCGTTGCATTCGATTCCTGTTTGTAATTGT 1293

 155 ...PheValProProArgValIle........................II 162
        |||          ||||||:::                         ::
1294 CCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAAT 1343

 162 eLysGluGlyAspValAspValSerAspSerAspAspGluAspAspAsnL 179
     ::::::|||    ||||||:::||||||    |||||||||    ::::::
1344 GAATAACGGTTTGGTTGATGCGACTGATTTTGATGACGAGCGTAATGGCT 1393

 179 euPro 180
        |||
1394 GGCCT 1398
```

seq name: /cgn2_6/ptodata/1/ina/6B COMB.seq:US-09-228-246-1

seq documentation_block:
; Sequence 1, Application US/09228246
; Patent No. 6245510
; GENERAL INFORMATION:
; APPLICANT: Staskawicz, B. S. et al.
; TITLE OF INVENTION: PRF Protein and Nucleic Acid Sequences: Compositions
; TITLE OF INVENTION: and Methods for Plant Pathogen Resistance
; FILE REFERENCE: 51700
; CURRENT APPLICATION NUMBER: US/09/228,246
; CURRENT FILING DATE: 1999-01-11
; EARLIER APPLICATION NUMBER: 08/680,427
; EARLIER FILING DATE: 1996-07-11
; EARLIER APPLICATION NUMBER: 08/310,912
; EARLIER FILING DATE: 1994-09-22
; EARLIER APPLICATION NUMBER: 08/227,360
; EARLIER FILING DATE: 1994-04-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 10968
; TYPE: DNA
; ORGANISM: Lycopersicon esculentum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3879)..(8186)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (8300)..(9466)
US-09-228-246-1

alignment_scores:

Quality:	102.00	Length:	699
Ratio:	0.360	Gaps:	29
Percent Similarity:	40.486	Percent Identity:	16.452

alignment_block:
US-09-932-678-2 x US-09-228-246-1 ..

Align seq 1/1 to: US-09-228-246-1 from: 1 to: 10968

```
 102 LeuThrLysAspPheGluGlnLeuIleSerIle...IleLeuArgLeuPr 117
     :::    |||    |||::::::    |||     :::    :::||||||
5484 ATATTGAAGAGGTTTCATGAATATATTCTTGTCAATCTGCTACGGAAGGA 5533

 117 oTrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLeuG 134
           |||     :::   |||:::::::::|||            |||
5534 TGAAACCAATTTGACATTTACTATTGCAGATGAGGTCAAAAAGTTTTATG 5583

 134 lyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMet 150
        :::|||:::            |||    ::::::::::|||
5584 ATGGGTTGTTGCTCATGGTTACATATCTTATTGAACCTCCA......... 5624

 151 IleAlaSerHisPheValProProArgValIleIleLysGluGlyAspVa 167
                     ||||||               |||:::::::|||::
5625 ................GTTCCTCACACTGAATGCAGGAAGCAAAATGATCT 5659

 167 lAspValSerAspSerAsp............................... 173
     :::::::        :::
5660 CTCAATGCGACATGAAGCTGTTGCAATTGAGGCGGAATCTGCTGTGTGTT 5709

 174 ..AspGluAspAspAsnLeuProAlaAsnPheAspThrCysHisArgAla 189
            :::|||||||:::        |||                |||
5710 TACATTATGACGATAATATGAATAACAAC.............AGTAGGGAG 5747

 190 LeuGlnIleIleAlaArgTyrValProSerThrProTrpPhe........ 203
     ::::::    :::   ::::::::::        |||    |||:::
5748 ATCAATCAGGTACTTCAGTTTTTGACTGTGACTTTCTGGCTTATCAAGTC 5797

 204 ...........LeuMetProIleLeuValGluLysPheProPheValArgL 217
                ||||||    :::|||    :::|||       :::    ::::
5798 TGAGGGTAACTTGATGGATCTACTGAAGCACAAATCCACTTTGGGAAATC 5847

 217 ysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSer 233
     ::               :::|||          ::::::|||:::
5848 AAGTTCTAGATCTGATTGAGAGTGCTCATGAAGAGCTTATT......... 5888

 234 ValTyrPheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLy 250
                   ||||||        :::::::::|||:::    :::||
5889 ............CTCCTTAGATCTATTCTCATGGATCTTCTTAGGAAAAA 5926

 250 sLeuLeuLysLeuAsp....................ValAsnAlaSerA 260
     ||||    :::||||||                  |||:::|||:::|
5927 GTTTACAGATTGGATGATCTCTTAATGCATGCTGAGGTGACTGCAAAAA 5976

 260 la................................................ 260
     ||
5977 GGTTAGCAATATTCAGTGGTTCTTGTTATGAATATTTCATGAACGGAAGC 6026

 261 ......................................GlnGlyIleGl 264
                                           |||    |||||
6027 AGCACTGAGAAAATGAGGCCCTTGTTATCTGATTTTCTGCAAGAGATTGA 6076

 264 uAspAlaGluGluThrAlaThrGlnThrCys.................. 274
     |:::::::::                 ::::::|||
6077 GTCTGTCAAGGTAGAGTTCAGAAATGTTTGCTTGCAAGTTCTGGATATAT 6126

 275 ........GlyGlyThrAspSerThrGluGlyLeuPheAsn......... 285
              :::    |||||:::      ||||||||    |||
6127 CACCTTTTTCCCTGACAGATGGA...GAAGGCCTTGTTAATTTCTTATTA 6173

 286 ................................................. 285

6174 AAAAACCAGGCCAAGGTGCCGAATGATGATGCTGTTTCTTCTGATGGAAG 6223

 286 .MetAspGluAspGluGluThrGluHis..................... 294
       :::::::::    :::::::||||||
6224 TTTAGAGGATGCAAGCAGCACTGAGAAAATGGGACTTCCATCTGATTTTC 6273

 295 ....................GluThrLysAlaGlyProGluArgLeuAsp 304
                         |||    |||   :::     :::        |||
6274 TCCGAGAGATTGAGTCTGTTGAGATAAAGGAGGCCAGAAAATTATATGAT 6323
```

```
 305 GlnMetVal.....HisProValAlaGluArgLeuAsp.....Il 316
     |||:::::::         |||       :::::::::   |||
6324 CAAGTTTTGGATGCAACACATTGTGAGACGAGTAAGACAGATGGAAAAAG 6373

 316 eLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspG 333
      :::::::::::::::|||:::   :::      :::
6374 CTTTATCAACATTATGTTAACCCAACAGGACAAGTTG.............C 6411

 333 lyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsn 349
           |||   |||:::::::::::  |||   :::::::::
6412 CGGACTATGATGCTGGTTCAGTCTCTTATCTTCTTAACCAAATATCAGTA 6461

 350 IlePheAspLysLeuLeu.....LeuProTh 358
     :::   ||||||||||||          :::
6462 GTTAAAGACAAACTATTGCACATTGGCTCTTTACTTGTAGATATTGTACA 6511

 358 rHisAlaSerCysHisValGln..... 365
      :::   :::   |||::::::
6512 GTACCGGAATATGCATATAGAACTTACAGATCTCGCTGAACGTGTTCAAG 6561

 365 ..... 365

6562 ATAAAAACTACATTTGTTTCTTCTCTGTCAAGGGTTATATTCCTGCTTGG 6611

 366 PhePheMetPheTyrLeuCysSerPheLys.....LeuGlyPheAlaGl 380
     ::::::   :::|||||   :::   |||      |||   |||:::||
6612 TATTACACACTATATCTCTCTGATGTCAAGCAATTGCTTAAGTTTGTTGA 6661

 380 uAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProA 397
     ||||   ::::::   :::      |||:::   |||   |||:::   :
6662 GGCAGAGGTAAAGATTATTTGTCTGAAAGTACCAGATTCTTCAAGTTATA 6711

 397 laIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArg 413
     :::::   :::      :::   ::::|||:::::   ||||||:::::
6712 GCTTCCCTAAGACAAATGGATTAGGATATCTCAATTGCTTTTTAGGCAAA 6761

 414 AlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuVa 430
        :::   :::          :::::|||   ||||||||||:::::
6762 TTGGAGGAGCTT.....TTACGTTCTAAGCTCGATTTGATAAT 6799

 430 lAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaP 447
     ::::      |||
6800 CGACTTAAAACAT..... 6812

 447 heCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaVal 463

6812 ..... 6812

 464 PheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLy 480
                          |||:::   |||      :::||
6813 .....CAGATTGAATCA.....GTCAA 6829

 480 sGluGlyLeuGlnTyrLeuGlnSerLeu.....AsnPheGluArgIle. 494
     |||||:|||||      |||:::|||:::      :::|||:::::::
6830 GGAGGGCTTATTGTGCCTAAGATCATTCATTGATCATTTTTCAGAAAGCT 6879

 495 .....ValMetSerGlnLeuAsnProLeu 502
                         ::::::::::::::::::   :::
6880 ATGATGAGCATGATGAAGCTTGTGGTCTTATAGCAAGAGTTTCTGTAATG 6929

 503 LysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLy 519
                     |||:::::::         |||   ::::::
6930 GCATACAAGGCTGAGTATGTCATTGACTCATGCTTGGCCTATTCTCATCC 6979

 519 s......TyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnA 534
           |||:::::::::::::          :::::|||   |||   :
6980 ACTCTGGTACAAAGTTCTTTGGATTTCTGAAGTTCTTGAG...AATATTA 7026

 534 rgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGln 550
     ::   :::      |||:::   :::|||:::          :::::::::
7027 AGCTTGTAAATAAAGTTGTTGGTGAGACATGTGAAAGAAGGAACATTGAA 7076

 551 IleCysThrAsnProLeuAspThrPhePheProPheAspProCysValLe 567
      :::   ::::::                                    ::
7077 GTTACTGTGCATGAA.....................................GT 7093

 567 uLysArgSerLysLysPheIleAspProIleTyrGlnValTrpGluAspM 584
      :   ::::::       ::::::   |||   ::::::::::::   :::
7094 TGCAAAGACTACCACTTATGTAGCACCATCTTTTTCAGCTTATACTCAAA 7143

 584 etSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle 600
       :::   ||||||:::::::   |||:::      :::
7144 GAGCAAACGAAGAAATGGAGGGTTTTCAGGATACAATA............. 7181

 601 ValGluAspGluAspAspAspPheLeuLysGlyGluValProGlnAsnAs 617
       ||||||      |||   |||   |||      ||:::   ||
7182 ......GATGAATTAAAGCATAAACTACTTGGAGGATCACCTGAGCTTGA 7225

 617 p......ThrValIleGlyIle..... 622
     |       ::::::::::|||:::
7226 TGTCATCTCAATCGTTGGCATGCCAGGATTGGGCAAGACTACACTAGCAA 7275

 623 .....ThrProSerSerPheAspThrHis 630
                              :::   |||   ||||||:::|||
7276 AGAAGATTTACAATGATCCAGAAGTCACCTCTCGCTTCGATGTCCAT 7322

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-680-327-1

seq_documentation_block:
; Sequence 1, Application US/08680327
; Patent No. 5859421
;  GENERAL INFORMATION:
;    APPLICANT:  Staskawicz, Brian S., Oldroyd, Giles Edward,
;    APPLICANT:  Salmeron, John M., Rommens, Caius
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR PLANT
;    TITLE OF INVENTION:  PATHOGEN RESISTANCE
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klarquist Sparkman Campbell Leigh &
;      ADDRESSEE:  Whinston
;      STREET:  One World Trade Center
;      STREET:  121 S.W. Salmon Street
;      STREET:  Suite 1600
;      CITY:  Portland
;      STATE:  Oregon
;      COUNTRY:  United States of America
;      ZIP:  97204
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Disk, 3-1/2 inch
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  MS DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/680,327
;      FILING DATE:  July 11, 1996
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/310,912
;      FILING DATE:  September 22, 1994
;      CLASSIFICATION:  800
;      APPLICATION NUMBER:  08/227,360
;      FILING DATE:  April 13, 1994
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dow, Alan. E.
;      REGISTRATION NUMBER:  35,124
;      REFERENCE/DOCKET NUMBER:  5151-45038
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (503) 226-7391
;      TELEFAX:  (503) 228-9446
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5475 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double stranded
;      TOPOLOGY:  linear
```

```
US-08-680-327-1

alignment scores:
              Quality:   101.00          Length:    699
                Ratio:    0.357            Gaps:     29
   Percent Similarity:   40.486   Percent Identity: 16.452

alignment block:
  US-09-932-678-2 x US-08-680-327-1 ..

  Align seq 1/1  to: US-08-680-327-1  from: 1  to: 5475

   102 LeuThrLysAspPheGluGlnLeuIleSerIle...IleLeuArgLeuPr 117
        :::     |||     |||:::::    |||     :::     :::||||||
  1606 ATATTGAAGAGGTTTCATGAATATATTCTTGTCAATCTGCTACGGAAGGA 1655

   117 oTrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLeuG 134
          |||     :::     |||::::::::::|||          |||
  1656 TGAAACCAATTTGACATTTACTATTGCAGATGAGGTCAAAAGTTTTATG 1705

   134 lyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMet 150
         :::|||:::          |||     :::::::::|||
  1706 AAGGTTGTTGCTCATGGTTACATATCTTATTGAACCTCCA......... 1746

   151 IleAlaSerHisPheValProProArgValIleIleLysGluGlyAspVa 167
                          ||||||          |||::::::|||::
  1747 ...............GTTCCTCACACTGAATGCAGGAAGCAAAATGATCT 1781

   167 lAspValSerAspSerAsp............................... 173
        ::::::          :::
  1782 CTCAATGCGACATGAAGCTGTTGCAATTGAGGCGGAATCTGCTGTGTGTT 1831

   174 ..AspGluAspAspAsnLeuProAlaAsnPheAspThrCysHisArgAla 189
             :::|||||||:::     |||          |||
  1832 TACATTATGAGGATAATATGAATAACAAC...........AGTAGGGAG 1869

   190 LeuGlnIleIleAlaArgTyrValProSerThrProTrpPhe........ 203
        ::::::     :::     :::::::::     |||     |||:::
  1870 ATCAATCAGGTACTTCAGTTTTGACTGTGACTTTCTGGCTTATCAAGTC 1919

   204 .........LeuMetProIleLeuValGluLysPheProPheValArgL 217
                 ||||||     :::|||     :::|||          :::     ::::
  1920 TGAGGTAACTTGATGGATCTACTGAAGCACAAATCCACTTTGGGAAATC 1969

   217 ysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSer 233
        ::          :::|||          :::::|||:::
  1970 AAGTTCTAGATCTGATTGAGAGTGCTCATGAAGAGCTTATT......... 2010

   234 ValTyrPheProThrLeuArgHisGlnIleLeuGluLeuIleIleGluLy 250
                    ||||||          :::::::::|||:::          :::||
  2011 ...........CTCCTTAGATCTATTCTCATGGATCTTCTTAGGAAAAA 2048

   250 sLeuLeuLysLeuAsp...................ValAsnAlaSerA 260
        ||||     :::||||||                    |||:::|||:::|
  2049 GCTTTACAGATTGGATGATCTCTTAATGCATCCTGAGGTGACTGCAAAAA 2098

   260 rg............................................. 260
        ||
  2099 GGTTAGCAATATTCAGTGGTTCTTGTTATGAATATTTCATGAACGGAAGC 2148

   261 .......................................GlnGlyIleGl 264
                                                   |||     |||||
  2149 AGCACTGAGAAAATGAGGCCCTTGTTATCTGATTTTCTGCAAGAGATTGA 2198

   264 nAspAlaGluGluThrAlaThrGluThrCys.................. 274
        |:::::::::          ::::::|||
  2199 GTCTGTCAAGGTAGAGTTCAGAAATGTTTGCTTGCAAGTTCTGGATATAT 2248

   275 ........GlyGlyThrAspSerThrGluGlyLeuPheAsn........ 285
                 :::     ||||||:::     ||||||||     |||
  2249 CACCTTTTTCCCTGACAGATGGA...GAAGGCCTTGTTAATTTCTTATTA 2295

   285 ................................................. 285

  2296 AAAAACCAGGCCAAGGTGCCGAATGATGATGCTGTTTCTTCTGATGGAAG 2345

   286 .MetAspGluAspGluGluThrGluHis..................... 294
         :::::::::     :::::||||||
  2346 TTTAGAGGATGCAAGCAGCACTGAGAAAATGGGACTTCCATCTGATTTTC 2395

   295 ...................GluThrLysAlaGlyProGluArgLeuAsp 304
                           |||     |||     :::     :::     |||
  2396 TCCGAGAGATTGAGTCTGTTGAGATAAAGGAGGCCAGAAAATTATATGAT 2445

   305 GlnMetVal.........HisProValAlaGluArgLeuAsp......Il 316
        |||::::::          |||          :::::::::     |||
  2446 CAAGTTTTGGATGCAACACATTGTGAGACGAGTAAGACAGATGGAAAAAG 2495

   316 eLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspG 333
         :::::::::::::::|||:::     :::          :::
  2496 CTTTATCAACATTATGTTAACCCAACAGGACAAGTTG...........C 2533

   333 lyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsn 349
                 |||     |||:::::::::     |||     :::::::::
  2534 CTGACTATGATGCTGGTTCAGTCTCTTATCTTCTTAACCAAATATCAGTA 2583

   350 IlePheAspLysLeuLeu.......................LeuProTh 358
        :::     ||||||||||||                        :::
  2584 GTTAAAGACAAATTATTGCCACATTGGCTCTTTACTTGTAGATATTGTACA 2633

   358 rHisAlaSerCysHisValGln........................... 365
         :::     :::     |||:::::
  2634 GTACCGGAATATGCATATAGAACTTACAGATCTCGCTGAACGTGTTCAAG 2683

   365 ................................................. 365

  2684 ATAAAAACTACATTTGTTTCTTCTCTGTCAAGGTTATATTCCTGCTTGG 2733

   366 ProPheMetPheTyrLeuCysSerPheLys......LeuGlyPheAlaGl 380
        ::::::     :::||||||     :::     |||          |||     |||:::||
  2734 TATTACACACTATATCTCTCTGATGTCAAGCAATTGCTTAAGTTTGTTGA 2783

   380 nAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProA 397
        ||||     ::::::     :::          |||:::     |||     |||:::     :
  2784 GGCAGAGGTAAAGATTATTTGTCTGAAAGTACCAGATTCTTCAAGTTATA 2833

   397 laIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArg 413
        :::::     :::          :::     :::|||:::::     ||||||:::::
  2834 GCTTCCCTAAGACAAATGGATTAGGATATCTCAATTGCTTTTTAGGCAAA 2883

   414 AlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuVa 430
              :::     :::               :::::|||     ||||||||||:::::
  2884 TTGGAGGAGCTT...........TTACGTTCTAAGCTCGATTTGATAAT 2921

   430 lAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaP 447
        ::::     |||
  2922 CGACTTAAAACAT..................................... 2934

   447 heCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaVal 463

  2934 ................................................. 2934

   464 PheTyrThrPheValPheArgHisLysGluLeuLeuSerGlyAsnLeuLy 480
                                          |||:::     |||          :::||
  2935 .........................CAGATTGAATCA......GTCAA 2951

   480 sGluGlyLeuGlnTyrLeuGlnSerLeu......AsnPheGluArgIle. 494
        |||||||||          |||:::|||:::          :::|||:::::
  2952 GGAGGGCTTATTGTGCCTAAGATCATTCATTGATCATTTTTCAGAAAGCT 3001

   495 ........................ValMetSerGlnLeuAsnProLeu 502
                                    :::::::::::::::::     :::
  3002 ATGATGAGCATGATGAAGCTTGTGGTCTTATAGCAACAGTTTCTGTAATG 3051

   503 LysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLy 519
```

```
                          |||::::::         |||   ::::::
    3052 GCATACAAGGCTGAGTATGTCATTGACTCATGCTTGGCCTATTCTCATCC 3101

     519 s......TyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnA 534
             |||:::::::::::::        ::::::|||   |||   :
    3102 ACTCTGGTACAAAGTTCTTTGGATTTCTGAAGTTCTTGAG...AATATTA 3148

     534 rqGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGln 550
         ::   :::      |||:::   :::|||:::          :::::::::
    3149 AGCTTGTAAATAAAGTTGTTGGTGAGACATGTGAAAGAAGGAACATTGAA 3198

     551 IleCysThrAsnProLeuAspThrPhePheProPheAspProCysValLe 567
         :::   ::::::                                     ::
    3199 GTTACTGTGCATGAA................................GT 3215

     567 uLysArgSerLysLysPheIleAspProIleTyrGlnValTrpGluAspM 584
         :   ::::::      ::::::   |||   ::::::::::::   :::
    3216 TGCAAAGACTACCACTTATGTAGCACCATCTTTTTCAGCTTATACTCAAA 3265

     584 etSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle 600
           :::   ||||||||:::::   |||:::      :::
    3266 GAGCAAACGAAGAAATGGAGGGTTTTCAGGATACAATA............ 3303

     601 ValGluAspGluAspAspAspPheLeuLysGlyGluValProGlnAsnAs 617
             ||||||      |||   |||   |||      |||:::   ||
    3304 ......GATGAATTAAAGGATAAACTACTTGGAGGATCACCTGAGCTTGA 3347

     617 p......ThrValIleGlyIle........................... 622
         |      ::::::::::|||:::
    3348 TGTCATCTCAATCGTTGGCATGCCAGGATTGGGCAAGACTACACTAGCAA 3397

     623 .......................ThrProSerSerPheAspThrHis 630
                                 :::   |||   ||||||:::|||
    3398 ACAAGATTTACAATGATCCAGAAGTCACCTCTCGCTTCGATGTCCAT 3444

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-228-246-3

seq_documentation_block:
; Sequence 3, Application US/09228246
; Patent No. 6245510
; GENERAL INFORMATION:
;  APPLICANT: Staskawicz, B. S. et al.
;  TITLE OF INVENTION: PRF Protein and Nucleic Acid Sequences:  Compositions
;  TITLE OF INVENTION: and Methods for Plant Pathogen Resistance
;  FILE REFERENCE: 51700
;  CURRENT APPLICATION NUMBER: US/09/228,246
;  CURRENT FILING DATE: 1999-01-11
;  EARLIER APPLICATION NUMBER: 08/680,327
;  EARLIER FILING DATE: 1996-07-11
;  EARLIER APPLICATION NUMBER: 08/310,912
;  EARLIER FILING DATE: 1994-09-22
;  EARLIER APPLICATION NUMBER: 08/227,360
;  EARLIER FILING DATE: 1994-04-13
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 5475
;   TYPE: DNA
;   ORGANISM: Lycopersicon esculentum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(5475)
US-09-228-246-3


alignment_scores:
            Quality:  101.00            Length:    699
              Ratio:  0.357              Gaps:     29
 Percent Similarity:  40.486   Percent Identity: 16.452

alignment_block:
 US-09-932-678-2 x US-09-228-246-3   ..

Align Seq 1/1  to: US-09-228-246-3  from: 1  to: 5475

     102 LeuThrLysAspPheGluGlnLeuIleSerIle...IleLeuArgLeuPr 117
         :::   |||   |||::::::   |||   :::   :::|||||||
    1606 ATATTGAAGAGGTTTCATGAATATATTCTTGTCAATCTGCTACGGAAGGA 1655

     117 oTrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLeuG 134
             |||   :::   |||:::::::::::|||         |||
    1656 TGAAACCAATTTGACATTTACTATTGCAGATGAGGTCAAAAAGTTTTATG 1705

     134 lyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMet 150
           :::|||:::      |||   :::::::::|||
    1706 AAGGGTTGTTGCTCATGGTTACATATCTTATTGAACCTCCA..... 1746

     151 IleAlaSerHisPheValProProArgValIleIleLysGluGlyAspVa 167
                     ||||||                |||::::::|||::
    1747 ..............GTTCCTCACACTGAATGCAGGAAGCAAAATGATCT 1781

     167 lAspValSerAspSerAsp............................... 174
         :::::::      :::
    1782 CTCAATGCGACATGAAGCTGTTGCAATTGAGGCGGAATCTGCTGTGTGTT 1831

     174 ..AspGluAspAspAsnLeuProAlaAsnPheAspThrCysHisArgAla 189
               :::|||||:::      |||              |||
    1832 TACATTATGAGGATAATATGAATAACAAC............AGTAGGGAG 1869

     190 LeuGlnIleIleAlaArgTyrValProSerThrProTrpPhe........ 204
         ::::::   :::   :::::::::       |||   |||:::
    1870 ATCAATCAGGTACTTCAGTTTTTGACTGTGACTTTCTGGCTTATCAAGTC 1919

     204 ..........LeuMetProIleLeuValGluLysProProPheValArgL 217
                 |||'||   :::|||   :::|||       :::   ::::
    1920 TGAGGGTAACTTGATGGATCTACTGAAGCACAAATCCACTTTGGGAAATC 1969

     217 ysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSer 233
         ::      :::|||       ::::::|||:::
    1970 AAGTTCTAGATCTGATTGAGAGTGCTCATGAAGAGCTTATT......... 2010

     234 ValTyrProProThrLeuArgHisGluIleLeuGluLeuIleIleGluLy 250
                     ||||||       ::::::::::|||:::   :::||
    2011 ............CTCCTTAGATCTATTCTCATGGATCTTCTTAGGAAAAA 2048

     250 sLeuLeuLysLeuAsp.....................ValAsnAlaSerA 260
         ||:|   :::|||||||                    |||:::|||:::|
    2049 GCTTTACAGATTGGATGATCTCTTAATGCATGCTGAGGTCACTGCAAAAA 2098

     260 ra.................................................. 260
         |.
    2099 GTTAGCAATATTCAGTGGTTCTTGTTATGAATATTTCATGAACGGAAGC 2148

     261 ....................................GlnGlyIleGl 264
                                            |||   |||||
    2149 AGCACTGAGAAAATGAGGCCCTTGTTATCTGATTTCTGCAAGAGATTGA 2198

     264 uAspAlaGluGluThrAlaThrGlnThrCys..................... 274
         |:::::::::            ::::::|||
    2199 GTCTGTCAAGGTAGAGTTCAGAAATGTTTGCTTGCAAGTTCTGGATATAT 2248

     275 ........GlyGlyThrAspSerThrGluGlyLeuPheAsn......... 285
                 :::   |||||||:::   |||||||||   |||
    2249 CACCTTTTTCCCTGACAGATGGA...GAAGGCCTTGTTAATTTCTTATTA 2295

     285 ..................................................... 285

    2296 AAAAACCAGGCCAAGGTGCCGAATGATGATGCTGTTTCTTCTGATGGAAG 2345

     286 .MetAspGluAspGluGluThrGluHis....................... 294
          :::::::::   ::::::||||||
    2346 TTTAGAGGATGCAAGCAGCACTGAGAAAATGGGACTTCCATCTGATTTTC 2395

     295 ....................GluThrLysAlaGlyProGluArgLeuAsp 304
                             |||   |||   :::   :::       |||
    2396 TCCGAGAGATTGAGTCTGTTGAGATAAAGGAGGCCAGAAAATTATATGAT 2445
```

```
 405 GlnMetVal.....HisProValAlaGluArgLeuAsp.....Il 416
     |||:::::::          |||      :::::::::   |||
2446 CAAGTTTGGATCCAACACATTGTGAGACGACTAAGACACATGGAAAAAG 2495

 416 eLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspG 333
     :::::::::::::::|||:::   :::       :::
2496 CTTTATCAACATTATGTTAACCCAACAGGACAAGTTG.........C 2533

 333 lyLysValAspAsnGlyLysPhrLysAspLeuTyrArqAspLeuIleAsn 349
          |||   |||:::::::::::   |||   :::::::::
2534 CGGACTATGATGCTGGTTCAGTCTCTTATCTTCTTAACCAAATATCAGTA 2583

 350 IlePheAspLysLeuLeu.....LeuProTh 358
     :::   ||||||||||||          :::
2584 GTTAAAGACAAATTATTGCACATTGGCTCTTTACTTGTAGATATTGTACA 2633

 358 rHisAlaSerCysHisValGln..... 365
     :::   :::   |||:::::::
2634 GTACCGGAATATGCATATAGAACTTACAGATCTCGCTGAACGTGTTCAAG 2683

 365 ..... 365

2684 ATAAAAACTACATTTGTTTCTTCTCTGTCAAGGGTTATATTCCTGCTTGG 2733

 366 PhePheMetPheTyrLeuCysSerPheLys.....LeuGlyPheAlaGl 380
     ::::::   :::||||||   :::   |||      |||   |||:::||
2734 TATTACACACTATATCTCTCTGATGTCAAGCAATTGCTTAAGTTTGTTGA 2783

 380 nAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProA 397
     ||||   ::::::   :::      |||:::   |||   |||:::   :
2784 GGCAGAGGTAAAGATTATTTGTCTGAAAGTACCAGATTCTTCAAGTTATA 2833

 397 laIleIleArqGluAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArq 413
     :::::   :::      :::   :::|||:::::::   |||||||::::::
2834 GCTTCCCTAAGACAAATGGATTAGGATATCTCAATTGCTTTTTAGGCAAA 2883

 414 AlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuVa 430
     :::   :::             :::::|||   ||||||||||:::::
2884 TTGGAGGACCTT.....TTACCTTCTAAGCTCGATTTGATAAT 2921

 430 lAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaP 447
     ::::      |||
2922 CGACTTAAAACAT..... 2934

 447 heCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaVal 463

2934 ..... 2934

 464 PheTyrThrPheValPheArqHisLysGlnLeuLeuSerGlyAsnLeuLy 480
                          |||:::   |||      :::||
2935 .....CAGATTGAATCA.....GTCAA 2951

 480 sGluGlyLeuGluTyrLeuGlnSerLeu.....AsnPheGluArqIle. 494
     ||||||||||      ||||:::|||:::      :::||||:::::
2952 GGAGGGCTTATTGTGCCTAAGATCATTCATTGATCATTTTTCAGAAAGCT 3001

 495 .....ValMetSerGlnLeuAsnProLeu 502
                          ::::::::::::::::::   :::
3002 ATGATGAGCATGATGAAGCTTGTGGTCTTATAGCAAGAGTTTCTGTAATG 3051

 503 LysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLy 519
                     |||::::::::      |||   ::::::
3052 GCATACAAGGCTGACTATGTCATTGACTCATGCTTGGCCTATTCTCATCC 3101

 519 s.....TyrGlnLeuValPheCysTyrThrIleIleGluArqAsnAsnA 534
          |||:::::::::::::          :::::::|||   |||   :
3102 ACTCTGGTACAAAGTTCTTTGGATTTCTGAAGTTCTTGAG...AATATTA 3148

 534 rqGlnMetLeuProValIleArqSerThrAlaGlyGlyAspSerValGln 550
     ::   :::      |||:::   :::|||:::          :::::::::
3149 AGCTTGTAAATAAAGTTGTTGGTGAGACATGTGAAAGAAGGAACATTGAA 3198

 551 IleCysThrAsnProLeuAspThrPhePhePhePheAspProCysValIle 567
     :::   ::::::                                     ::
3199 GTTACTGTGCATCAA.....GT 3215

 567 GlysArqSerLysLysPheIleAspProIleTyrGlnValTrpGluAspM 584
     :   ::::::   ::::::   |||   ::::::::::::   :::
3216 TGCAAAGACTACCACTTATGTAGCACCATCTTTTTCAGCTTATACTCAAA 3265

 584 etSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle 600
         :::   ||||||:::::   |||:::      :::
3266 CAGCAAACGAAGAAATGGAGGGTTTCAGGATACAATA..... 3303

 601 ValGluAspGluAspAspAspAspPheLeuLysGlyGluValProGlnAsnAs 617
          ||||||      |||   |||   |||      |||:::   ||
3304 .....GATGAATTAAAGGATAAACTACTTGGAGGATCACCTGAGCTTGA 3347

 617 p.....ThrValIleGlyIle..... 622
     |      ::::::::::|||:::
3348 TGTCATCTCAATCGTTGGCATGCCAGGATTGGGCAAGACTACACTAGCAA 3397

 623 .....ThrProSerSerPheAspThrHis 630
                          :::   |||   |||||||:::|||
3398 AGAAGATTTACAAIGATCCAGAAGTCACCTCTCGCTTCGATGTCCAT 3444

seq_name: zeqn2_6/prodata/1/ina/5B_COMB.seq:US-09-080-897-3

seq documentation block:
; Sequence 3, Application US/09080897
; Patent No. 5985574
;  GENERAL INFORMATION:
;    APPLICANT:  King, Mary-Claire
;    APPLICANT:  Lynch, Eric D.
;    APPLICANT:  Lee, Ming
;    APPLICANT:  Morrow, Jan E.
;    APPLICANT:  Welcsh, Piri L.
;    APPLICANT:  Leon, Pedro E.
;    TITLE OF INVENTION:  Modulators of Actin
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  75 DENISE DRIVE
;      CITY:  HILLSBOROUGH
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94010
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/080,897
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UW97-001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 343-4341
;      TELEFAX:  (650) 343-4342
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4478 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-080-897-3


alignment_scores:
            Quality:  100.50          Length:    572
```

```
            Ratio:  0.387                Gaps:     25
 Percent Similarity:  45.455   Percent Identity: 19.755

alignment_block:
 US-09-932-678-2 x US-09-080-897-3   ..

 Align seq 1/1  to: US-09-080-897-3  from: 1  to: 4378

       1 MetAlaAlaProLeuLeuHisThrArgLeuProGly......AspAlaAl 15
         |||:::::::      |||||||||          |||       :::
     436 ATGGTGTCGCAATATCTGCACACTTCCAAGGCTGGCATGAACCAGAAAGA 485

      15 aAlaSerSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleS 32
         :::|||   |||||||:::      :::         |||::::::::::
     486 GAGCTCTAGGTCTGCCATGATGTACATCCAGGAGCTGAGGTCGGGCTTGC 535

      32 erAsnMet..................ArgAlaLeuGluAsnAspPhePhe 42
         :::|||                    :::::::|||:::   ::::::
     536 GGGATATGCACCTGCTTAGCTGCCTTGAGTCCCTTCGAGTCTCTCTCAAC 585

      43 AsnSerProProArgLysThrValArgPheGly............GlyTh 55
         |||:::|||       :::      ||||||               :::::
     586 AATAACCCTGTCAGTTGGGTGCAGACATTTGGTGCTGAGGGCCTAGCCTC 635

      55 rValThrGluValLeu......LeuLysTyrLysLysGlyGluThr.... 68
         ::::   ::::::|||      |||      :::|||   ||||||
     636 CTTATTGGACATCCTCAAACGACTCCATGATGAGAAAGAGGAGACTTCTG 685

      69 .................AsnAspPheGluLeuLeuLysAsnGlnLeuLeu 79
                         |||:::   |||::::::::::          |||
     686 GAAACTACGACAGCCCAAACCAGCATGAGATTATCCGC......TGTTTG 729

      80 AspProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheAr 96
                 :::::::::       |||:::   :::|||||   ::
     730 AAGGCTTTCATGAACAACAAGTTTGGAATCAAAACTATGTTGGAGACGGA 779

      96 gSerSerIleMetTyrLeuThrLys..................... 104
         :::::::|||:::   |||::::::
     780 AGAAGGAATCCTACTGCTGGTCAGAGCCATGGATCCTGCTGTTCCCAATA 829

     105 ........AspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrp 118
                 |||       :::|||:::|||   :::      ||||||
     830 TGATGATTGATGCAGCAAAGCTGCTGTCTGCCCTCTGTATCCTGCCGCAG 879

     119 .........LeuAsnArg............................Se 122
                 :::|||:::                            ::
     880 CCGGAGGACATGAATGAACGAGTTCTAGAGGCAATGACAGAGAGAGCTGA 929

     122 rGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerA 139
         ::::      ||||||:::::      :::|||   :::|||   |||:
     930 AATGGATGAGGTCGAACGCTTCCAGCCACTTCTGGACGGATTAAAAAGTG 979

     139 laGlnThrValPheLeuArg...ProCysLeuSerMetIleAlaSerHis 154
         ::   ::::::   |||:::      |||||||:::::|||   :::
     980 GGACCTCTATTGCCCTCAAAGTCGGATGCCTACAGCTCATCAATGCTCTC 1029

     155 PheValProProArg................................ 159
         :::::::|||   :::
    1030 ATCACTCCAGCTGAGGAACTGGACTTCCGAGTTCACATCCGAAGTGAGCT 1079

     159 ..................................................... 159

    1080 GATGCGCCTGGGGCTGCATCAGGTGTTGCAGGAGCTTCGAGAGATTGAAA 1129

     160 ........................ValIleIleLysGluGlyAspVal 167
                                 |||:::   ::::::||||||
    1130 ATGAAGATATGAAACTACAGCTGTGCGTGTTTGATGAACAAGGCGATGAA 1179

     168 AspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPheAs 184
         |||      |||         :::   |||:::::      :::::::||
    1180 GATTTCTTTGATCTGAAGGGACGGCTGGATGATATCGGCATGGAGATGGA 1229

     184 pPhrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr. 200

         |        ::::::::: ||||||||   :::   |||   ::::::
    1230 TCACTTTGGTGAAGTTTTTCAGATTATTTTAAACACAGTGAAAGATTCAA 1279

     201 ........ProTrpPheLeuMetProIleLeuValGluLysPheProPhe 214
                 |||   ||||||          :::::::::   :::   :::
    1280 AGGCAGAGCCACACTTCCTGTCT......ATCTTGCAGCATCTCCTGTTG 1323

     215 ValArgLysSerGluArgThrLeu.......................Gl 223
         ||||||::::::   ::::::                          ||
    1324 GTCCGAAATGATTATGAAGCCAGGCCACAGTACTATAAACTGATTGAAGA 1373

     223 uCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuA 240
         ||||       :::   :::|||:::   :::           |||   ::::
    1374 ATGTGTTTCTCAGATAGTTCTACACAAAAATGGAACTGATCCTGACTTCA 1423

     240 rgHisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspVal 256
         ::   :::   |||::::::   ||||||:::|||       :::|||
    1424 AGTGCCGACACCTGCAGATTGATATTGAGAGATTG......GTTGATCAA 1467

     257 AsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThrGlnTh 273
           ::::::    :::|||   :::|||      ||||||:::
    1468 ATGATTGATAAAACAAAGGTGGAAAAATCTGAGGCCAAAGCTACAGAGCT 1517

     274 r*ysGlyGlyThrAspSer.......................ThrGluG 282
                    ||||||
    1518 GGAAAAAAAGTTGGATTCAGAATTAACAGCGCGGCACGAGTTACAAGTAG 1567

     282 lyLeuPheAsnMetAspGluAsp...GluGluThrGluHisGluThrLys 297
         :::   :::|||:::::|||   |||:::      ::::::   :::
    1568 AAATGAAAAAGATGGAAAATGACTTTGAGCAGAAACTTCAGGATCTTCAA 1617

     298 AlaGlyProGluArgLeuAspGlnMetValHisProValAlaGluArgLe 314
         :::      :::   |||||||:::      :::   ::::::   :::
    1618 GGAGAAAAGGATGGCTTGGATTCTGAAAAGCAGCAGATCACTGCACAGAA 1667

     314 uAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrV 331
         :::   |||   :::   |||   :::          :::|||:::   :
    1668 ACAAGACCTGGAGGCAGAGGTGTCCAAGCTGACAGGAGAGGTTGCCAAGC 1717

     331 alAspGlyLysValAspAsnGlyLysThr.LysAspLeuTyrArgAspLe 347
         :::::   :::::::::::::::::|||::: |||   |||          ||
    1718 TGTCAAAAGAACTAGAAGATGCCAAGAATGAAATGGCTTCTCTCTCTGCT 1767

     347 u.............IleAsnIlePheAspLysLeuLeuLeuProThrHisA 360
         |            :::   ::::::      ||||||:::|||
    1768 GTGGTTGTTGCACCTTCTGTTTCTAGCAGTGCTGCTGTTCCCCCTGCCCC 1817

     360 laSerCysHisValGlnPhe...PheMetPheTyrLeuCysSerPheLys 375
           |||   |||   :::   ::::::|||:::|||
    1818 TCCTCTGCCTGGTGACTCTGGCACTGTTATTCCACCTCCCCCACCCCCAC 1867

     376 LeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAs 392
         |||   |||   |||::::::   :::|||
    1868 CTCCTCTTCCTGGAGGTGTGGTCCCACCAT..................... 1897

     392 pProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlyS 409

    1897 ..................................................... 1897

     409 erPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCys 425
                                      |||||||::            |||
    1898 ........................CCCCTCCTC..........TGC 1909

     426 LeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSe 442
         |||:::::|||||||:::   ||||||::   |||
    1910 CTCCAGGTACTTGTATCC...CTCCACCTCCTCCTT.............. 1942

     442 rGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerA 459
                 :::   :::|||   ::::::   ||||::      :
    1943 .............TACCTGGAGGTGCTTGTATAC...CCCCTCCCCCCCA 1976

     459 laCysGlnAlaVal 463
         ::|||   ||||||
```

```
     1977 GTTGCCTGGCAGTG 1990

Seq name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-323-735-3

Seq documentation block:
; Sequence 3, Application US/09323735
; Patent No. 6197942
;  GENERAL INFORMATION:
;    APPLICANT:  King, Mary Claire
;    APPLICANT:  Lynch, Eric D.
;    APPLICANT:  Lee, Ming
;    APPLICANT:  Morrow, Jan E.
;    APPLICANT:  Welcsh, Piri L.
;    APPLICANT:  Leon, Pedro E.
;    TITLE OF INVENTION:  Modulators of Actin
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  75 DENISE DRIVE
;      CITY:  HILLSBOROUGH
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94010
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/323,735
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/080,897
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UW97-001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 343-4341
;      TELEFAX:  (650) 343-4342
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-323-735-3


alignment_scores:
             Quality:   100.50               Length:     572
               Ratio:   0.387                  Gaps:      25
 Percent Similarity:   45.455    Percent Identity:  19.755

alignment_block:
 US-09-932-678-2 x US-09-323-735-3      ..

 Align seq 1/1  to: US-09-323-735-3  from: 1  to: 4378

      1 MetAlaAlaProLeuLeuHisThrArgLeuProGly......AspAlaAl 15
        |||:::::::        |||||||||            |||      :::
    436 ATGGTGTCGCAATATCTGCACACTTCCAAGCTGGCATGAACCAGAAAGA 485

     15 aAlaSerSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleS 32
          :::|||    ||||||:::          :::           |||:::|||:::
    486 GAGCTCTAGGTCTGCCATGATGTACATCCAGGAGCTGAGGTCGGGCTTGC 535

     32 erAsnMet......................ArgAlaLeuGluAsnAspPhePhe 42
           :::|||                   ::::::|||:::    ::::::
    536 GGGATATGCACCTGCTTAGCTGCCTTGAGTCCCTTCGAGTCTCTCTCAAC 585

     43 AsnSerProProArgLysThrValArgPheGly............GlyTh 55
        |||::::|||           :::       ||||||             :::::
    586 AATAACCCTGTCAGTTGGGTGCAGACATTTGGTGCTGAGGGCCTAGCCTC 635

     55 rValThrGluValLeu......LeuLysTyrLysLysGlyGluThr.... 68
        ::::     ::::::|||        |||        :::|||    ||||||
    636 CTTATTGGACATCCTCAAACGACTCCATGATGAGAAAGAGGAGACTTCTG 685

     69 ...... ..........AsnAspPheGluLeuLeuLysAsnGlnLeuLeu 79
                         |||:::    |||::::::::::          |||
    686 GAAACTACGACAGCCGAAACCAGCATGAGATTATCCGC......TGTTTG 729

     80 AspProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheAr 96
                 :::::::::             |||:::    :::||||||     ::
    730 AAGGCTTTCATGAACAACAAGTTTGGAATCAAAACTATGTTGGAGACGGA 779

     96 qSerSerIleMetTyrLeuThrLys......................... 104
        :::::::|||:::     |||::::::
    780 AGAAGGAATCCTACTGCTGGTCAGAGCCATGGATCCTGCTGTTCCCAATA 829

    105 ........AspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrp 118
                 |||         :::|||:::|||     :::       ||||||
    830 TGATGATTGATGCAGCAAAGCTGCTGTCTGCCCTCTGTATCCTGCCGCAG 879

    119 .........LeuAsnArg.............................Se 122
                  :::|||:::                                 ::
    880 CTGGAGGACATGAATGAACGAGTTCTAGAGGCAATGACAGAGAGAGCTGA 929

    122 rGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerA 139
        ::::         ||||||:::::         :::|||     :::|||     |||:
    930 AATGGATGAGGTCGAACGCTTCCAGCCACTTCTGGACGGATTAAAAAGTG 979

    139 laGlnThrValPheLeuArg...ProCysLeuSerMetIleAlaSerHis 154
        ::     ::::::      |||:::          ||||||::::::|||      :::
    980 GGACCTCTATTGCCGTCAAAGTGGGATGCCTACAGCTCATCAATGCTCTC 1029

    155 PheValProProArg................................... 159
        ::::::|||     :::
   1030 ATCACTCCAGCTGAGGAACTGGACTTCCGAGTTCACATCCGAAGTGAGCT 1079

    159 .................................................. 159

   1080 GATGCGCCTGGGGCTGCATCAGGTGTTGCAGGAGCTTCGAGAGATTGAAA 1129

    160 ..........................ValIleIleLysGluGlyAspVal 167
                                   |||:::    ::::::||||||
   1130 ATGAAGATATGAAAGTACAGCTGTGCGTGTTTGATGAACAAGGGGATGAA 1179

    168 AspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPheAs 184
        |||        |||          :::    |||::::::         :::::::||
   1180 GATTTCTTTGATCTGAAGGGACGGCTGGATGATATCCGCATGGAGATGGA 1229

    184 pThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr. 200
        |              ::::::::::||||||||     :::     |||     ::::::
   1230 TGACTTTGGTGAAGTTTTTCAGATTATTTTAAACACAGTGAAAGATTCAA 1279

    201 ........ProTrpPheLeuMetProIleLeuValGluLysPheProPhe 214
                 |||    ||||||             :::::::::    :::     :::
   1280 AGGCAGAGCCACACTTCCTGTCT......ATCTTGCAGCATCTCCTGTTG 1323

    215 ValArgLysSerGluArgThrLeu.........................Gl 223
        ||||||::::::    ::::::                                  ||
   1324 GTCCGAAATGATTATGAAGCCAGGCCACAGTACTATAAACTGATTGAAGA 1373

    223 nCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuA 240
        ||||         :::     :::|||:::     :::                |||     ::::
   1374 ATGTGTTTCTCAGATAGTTCTACACAAAAATGGAACTGATCCTGACTTCA 1423

    240 rgHisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspVal 256
        ::     :::     |||:::::     ||||||:::|||          :::|||
   1424 AGTGCCGACACCTGCAGATTGATATTGAGAGATTG......GTTGATCAA 1467
```

```
     257 AsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThrGlnTh 273
         :::::      ::::::   :::::      ::::::::::
    1468 ATGATTGATAAAACAAAGGTGGAAAAATCTGAGGCCAAAGCTACAGAGCT 1517

     273 rCysGlyGlyThrAspSer.....ThrGluG 282
                 ::::::
    1518 GCAAAAAAAGTTGGATTCAGAATTAACAGCGCGGCACGAGTTACAAGTAG 1567

     282 lyLeuPheAsnMetAspGluAsp...GluGluThrGluHisGluThrLys 297
           :::   ::::::::::::::::   ::::::      ::::::   :::
    1568 AAATGAAAAAGATGGAAAATGACTTTGAGCAGAAACTTCAGGATCTTCAA 1617

     298 AlaGlyProGluArgLeuAspGlnMetValHisProValAlaGluArgLe 314
         :::      :::   :::::::::      :::   ::::::   :::
    1618 GGAGAAAAGGATGCCTTGGATTCTGAAAAGCAGCAGATCACTGCACAGAA 1667

     314 uAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrV 331
          :::   :::   :::   :::   :::          :::::::::  :
    1668 ACAAGACCTGGAGGCAGAGGTGTCCAAGCTGACAGGAGAGGTTGCCAAGC 1717

     331 alAspGlyLysValAspAsnGlyLysThr.LysAspLeuTyrArgAspLe 347
         :::::   ::::::::::::::::::::::: :::   :::         ::
    1718 TGTCAAAAGAACTAGAAGATGCCAAGAATGAAATGGCTTCTCTCTCTGCT 1767

     347 u.....IleAsnIlePheAspLysLeuLeuProThrHisA 360
         :            :::   ::::::      :::::::::::
    1768 GTGGTTGTTGCACCTTCTGTTTCTAGCAGTGCTGCTGTTCCCCCTGCCCC 1817

     360 laSerCysHisValGlnPhe...PheMetPheTyrLeuCysSerPheLys 375
           :::   :::   :::   :::::::::::::::
    1818 TCCTCTGCCTGGTGACTCTGGCACTGTTATTCCACCTCCCCCACCCCCAC 1867

     376 LeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAs 392
         :::   :::   :::::::::   ::::::
    1868 CTCCTCTTCCTGGAGGTGTGGTCCCACCAT..... 1897

     392 pProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlyS 409

    1897 ..... 1897

     409 erPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCys 425
                                 :::::::::              :::
    1898 .........................CCCCTCCTC............TGC 1909

     426 LeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSe 442
         :::::::::::::::::   :::::::::   :::
    1910 CTCCAGGTACTTGTATCC...CTCCACCTCCTCCTT..... 1942

     442 rGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerA 459
                   :::   ::::::   ::::::   :::::::      :
    1943 .............TACCTGGAGGTGCTTGTATAC...CCCCTCCCCCCCA 1976

     459 laCysGlnAlaVal 463
         ::::::   ::::::
    1977 GTTGCCTGGCAGTG 1990

seq_name: /cqn2_6/ptodata/1/ina/6A_COMB.seq:US-08-899-595-2

seq_documentation_block:
; Sequence 2, Application US/08899595
; Patent No. 6111072
;  GENERAL INFORMATION:
;    APPLICANT:  Narumiya, Shuh
;    APPLICANT:  Takahashi, No. 6111072uaki
;    TITLE OF INVENTION:  RHO TARGET PROTEIN HUMAN MDIA AND GENE
;    TITLE OF INVENTION:  ENCODING SAME
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patentin Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/899,595
;      FILING DATE:  24-JUL-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 8-242701
;      FILING DATE:  26-AUG-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 9-90170
;      FILING DATE:  25-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stephen A. Bent
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  049441/0112
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4399 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Mouse
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  94..3858
US-08-899-595-2


alignment_scores:
            Quality:  100.50              Length:    572
              Ratio:  0.387                 Gaps:     25
 Percent Similarity:  45.455   Percent Identity: 19.755

alignment_block:
 US-09-932-678-2 x US-08-899-595-2   ..

 Align seq 1/1  to: US-08-899-595-2  from: 1  to: 4399

       1 MetAlaAlaProLeuLeuHisThrArgLeuProGly......AspAlaAl 15
         ||:::::::      ||||||||||      |||      :::
     430 ATGGTGTCGCAATATCTGCACACTTCCAAGGCTGGCATGAACCAGAAAGA 485

      15 aAlaSerSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleS 32
         ::::|||   |||||||:::      :::          |||::|||:::
     486 GAGCTCTAGGTCTGCCATGATGTACATCCAGGAGCTGAGGTCGGGCTTGC 535

      33 erAsnMet.....ArgAlaLeuGluAsnAspPhePhe 42
           :::|||                 ::::::|||:::   ::::::
     536 GTGATATGCACCTGCTTAGCTGCCTTGAGTCCCTTCGAGTCTCTCTCAAC 585

      43 AsnSerProProArgLysThrValArgPheGly.....GlyTh 55
         |||::|||          :::      ||||||          :::::
     586 AATAACCCTGTCAGTTGGGTGCAGACATTTGGTGCTGAGGGCCTAGCCTC 635

      55 rValThrGluValLeu.....LeuLysTyrLysLysGlyGluThr.... 68
         ::::   ::::::|||      |||      :::|||   ||||||
     636 CTTATTGGACATCCTCAAACGACTCCATGATGAGAAAGAGGAGACTTCTG 685

      69 .............AsnAspPheGluLeuLeuLysAsnGlnLeuLeu 79
                      |||:::   |||:::::::::          |||
     686 GAAACTACGACAGCCGAAACCAGCATGAGATTATCCGC......TGTTG 729
```

```
  80 AspProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheAr 96
               ::::::::::           |||:::    :::||||||     ::
 730 AAGGCTTTCATGAACAACAAGTTTGGAATCAAAACTATGTTGGACACGGA 779

  96 gSerSerIleMetTyrLeuThrLys......................... 104
     :::::::|||:::     |||::::::
 780 AGAAGGAATCCTACTGCTGGTCAGAGCCATGGATCCTGCTGTTCCCAATA 829

 105 ........AspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrp 118
             |||        :::|||:::|||     :::        ||||||
 830 TGATGATTGATGCAGCAAAGCTGCTGTCTGCCCTCTGTATCCTGCCGCAG 879

 119 .........LeuAsnArg............................Se 122
          :::|||:::                                    ::
 880 CCGGAGGACATGAATGAACGAGTTCTAGAGGCAATGACAGAGAGAGCTGA 929

 122 rGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerA 139
     ::::         ||||||::::::          :::|||      :::|||    |||:
 930 AATGGATGAGGTCGAACGCTTCCAGCCACTTCTGGACGGATTAAAAAGTG 979

 139 laGlnThrValPheLeuArg...ProCysLeuSerMetIleAlaSerHis 154
     ::   ::::::    |||:::          ||||||:::::|||     :::
 980 GGACCTCTATTGCCCTCAAAGTGGGATGCCTACAGCTCATCAATGCTCTC 1029

 155 PheValProProArg................................... 159
     ::::::|||    :::
1030 ATCACTCCAGCTGAGGAACTGGACTTCCGAGTTCACATCCGAAGTGAGCT 1079

 159 .................................................. 159

1080 GATGCGCCTGGGGCTGCATCAGGTGTTGCAGGAGCTTCGAGAGATTGAAA 1129

 166 ..........................ValIleIleLysGluGlyAspVal 167
                               |||:::     ::::::||||||
1130 ATGAAGATATGAAAGTACAGCTGTGCGTGTTTGATGAACAAGGGGATGAA 1179

 168 AspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPheAs 184
     |||         |||          :::    |||::::::        ::::::||
1180 GATTTCTTTGATCTGAAGGCACGGCTGGATGATATCCGCATGGAGATGGA 1229

 184 pThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr. 200
     |                ::::::::|||||||||     :::     |||     ::::::
1230 TGACTTTGGTGAAGTTTTTCAGATTATTTTAAACACAGTGAAAGATTCAA 1279

 201 ........ProTrpPheLeuMetProIleLeuValGluLysPheProPhe 214
             |||    ||||||           :::::::::    :::    :::
1280 AGGCAGAGCCACACTTCCTGTCT......ATCTTGCAGCATCTCCTGTTG 1323

 215 ValArgLysSerGluArgThrLeu.......................Gl 223
     ||||||::::::     ::::::                                ||
1324 GTCCGAAATGATTATGAAGCCAGGCCACAGTACTATAAACTGATTGAAGA 1373

 223 uCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLeuA 240
     ||||           :::     :::|||:::     :::           |||     ::::
1374 ATGTGTTTCTCAGATAGTTCTACACAAAAATGGAACTGATCCTGACTTCA 1423

 240 rgHisGluIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspVal 256
     ::    :::      |||::::::      ||||||:::|||          :::|||
1424 ACTGCCGACACCTGCAGATTGATATTGAGAGATTG......GTTCATCAA 1467

 257 AsnAlaSerArgGlnGlyIleGluAspAlaGlnGluThrAlaThrGluTh 273
             ::::::          :::|||     :::|||          ||||||:::
1468 ATGATTGATAAAACAAAGGTGGAAAAATCTGAGGCCAAAGCTACAGAGCT 1517

 273 rCysGlyGlyThrAspSer.......................ThrGluG 282
                   ||||||
1518 GGAAAAAAAGTTGGATTCAGAATTAACAGCGCGGCACGAGTTACAACTAG 1567

 282 lyLeuPheAsnMetAspGlnAsp...GluGluThrGluHisGluThrLys 297
       :::      :::|||:::::::|||      |||:::         ::::::    :::
1568 AAATGAAAAAGATGGAAAATGACTTTGAGCAGAAACTTCAGGATCTTCAA 1617

 298 AlaGlyProGluArgLeuAspGluMetValHisProValAlaGluArgLe 314
       :::       :::    ||||||:::          :::    ::::::    :::
1618 GGAGAAAACGGATGCCTTGGATTCTGAAAAGCAGCAGATCACTGCACAGAA 1667

 314 uAspIleLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrV 331
      :::     |||     :::     |||     :::               :::|||:::    :
1668 ACAAGACCTGGAGGCAGAGGTGTCCAAGCTGACAGGAGAGGTTGCCAAGC 1717

 331 alAspGlyLysValAspAsnGlyLysThr.LysAspLeuTyrArgAspLe 347
     :::::     ::::::::::::::::|||::: |||     |||              ||
1718 TGTCAAAAGAACTAGAAGATGCCAAGAATGAAATGGCTTCTCTCTCTGCT 1767

 347 u............IleAsnIlePheAspLysLeuLeuLeuProThrHisA 360
     |              :::     ::::::        ||||||:::|||
1768 GTGGTTGTTGCACCTTCTGTTTCTAGCAGTGCTGCTGTTCCCCCTGCCCC 1817

 360 laSerCysHisValGlnPhe...PheMetPheTyrLeuCysSerPheLys 375
         |||     |||    :::    ::::::|||:::|||
1818 TCCTCTGCCTGGTGACTCTGGCACTGTTATTCCACCTCCCCCACCCCCAC 1867

 376 LeuGlyPheAlaGluAlaPheLeuGlnHisLeuTrpLysLysLeuGlnAs 392
     |||     |||     |||::::::     :::|||
1868 CTCCTCTTCCTGGAGGTGTGGTCCCACCAT.................... 1897

 392 pProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlyS 409

1897 ................................................. 1897

 409 erPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerCys 425
                              ||||||:::               |||
1898 .........................CCCCTCCTC............TGC 1909

 426 LeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSe 442
     |||::::::||||||:::     ||||||:::     |||
1910 CTCCAGGTACTTGTATCC...CTCCACCTCCTCCTT.............. 1942

 442 rGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerA 459
                      :::     :::|||     ::::::     |||:::     :
1943 .............TACCTGGAGGTGCTTGTATAC...CCCCTCCCCCCA 1976

 459 laCysGlnAlaVal 463
     ::|||    ||||||
1977 GTTGCCTGGCAGTG 1990
```

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-973-462-2

seq_documentation_block:
; Sequence 2, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 5361
;   TYPE: DNA
;   ORGANISM: P. falciparum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(5361)
US-08-973-462-2

alignment_scores:
Quality: 99.50 Length: 559

```
              Ratio:   0.381                Gaps:      23
  Percent Similarity:  46.691   Percent Identity: 19.499

alignment_block:
 US-09-932-678-2 x US-08-973-462-2   ..

 Align seq 1/1  to: US-08-973-462-2  from: 1  to: 5361

      31 IleSerAsnMetArgAlaLeuGluAsnAspPhePheAsnSerProProAr 47
         ::::|||::::::::::   |||||:::|||::::::::::::      ::
    3949 GTATCTGATTTAAAAGATCTTGAAGAAGATATATTAAAAGAAGTAAAAGA 3998

      47 g.............................................LysT 49
         :                                             ||||
    3999 AATCAAAGAACTTGAAAGTGAAATTTTAGAAGATTATAAAGAATTAAAAA 4048

      49 hrValArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLys 65
         ||:::::              |||:::::::|||    :::    |||:::
    4049 CTATTGAA...............ACAGATATTTTAGAAGAGAAAAAAGAA 4083

      66 GlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAs 82
             |||     :::    ||||||   ::::::::::::      ::
    4084 ATAGAAAAACATCATTTTGAAAAATTCGAAGAAGAAGCT......GAAGA 4127

      82 pIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSer.... 97
         :|||||||||    :::       :::    |||    |||      |||
    4128 AATAAAAGATCTTGAA...GCAGATATATTAAAAGAAGTATCTTCATTAG 4174

      98 .........................SerIleMetTyrLeuThrLysAsp 105
                                  ::::::      |||    ::::::
    4175 AAGTTGAAGAAGAAAAAAAATTAGAAGAAGTACACGAATTAAAAGAAGAG 4224

     106 PheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgSe 122
             |||::::::||||||           ::::::      |||:::::::::
    4225 GTAGAACATATAATAAGTGGTGATGCGCATATAAAAGCTTTGGAAGAAGA 4274

     122 rGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerA 139
         ::::        :::||||||         :::     |||:::::
    4275 TGAT......TTAGAAGAAGTAGATGATTTAAAAGGAAGTATA....... 4311

     139 laGlnThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPhe 155
                                  |||::::|||:::
    4312 .........................TTAGACATGTTA............ 4323

     156 ValProProArgValIleIleLysGluGlyAspValAspValSerAspSe 172
                                 :::||||||::::::::::::::|||
    4324 ........................AAGGGAGATATGGAATTAGGGGATAT 4349

     172 rAspAspGluAsp..........................AspAsnLeuP 180
           |||    |||:::                         :::::::::
    4350 GGATAAGGAAAGTTTAGAAGATGTAACAACAAAACTTGGAGAAACAGTTG 4399

     180 roAlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyr 196
            :::          ||||:::        ::::::  :::       :::
    4400 AATCCTTAAAAGATGTTTTATCTAGTGCATTAGGCATGGATGAAGAACAA 4449

     197 ValProSerThrProTrpPheLeuMetProIleLeuValGluLysPhePr 213
         :::    :::                   |||    |||    |||    :::
    4450 ATGAAAACAAGAAAAAAAGCTCAAAGACCTAAGTTGGAAGAAGTATTATT 4499

     213 oPheValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeuL 230
                :::    ::::|||:::       :::       :::    :::    :
    4500 AAAAGAAGAGGTTAAAGAAGAACCAAAGAAAAAAAATAACAAAAAAGAAAG 4549

     230 euArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLeu 246
         ::|||:::::::::                   :::    ||||||:::|||:::
    4550 TAAGGTTTGATATTAAGGATAAGGAACCAAAAGATGAAATAGTAGAAGTT 4599

     247 .........................IleIleGluLysLeuLeuLysLe 254
                                  :::|||:::     ::::::
    4600 GAAATGAAAGATGAAGATATAGAAGAAGATGTAGAAGATATAGAAGA 4649

     254 uAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaT 271
         |||::::::    ::::::          ::::::|||||||    ::::|||
    4650 AGATATAGAAGAAGATAAA......GTTGAAGATATAGATGAAGATATAG 4693

     271 hrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAsp 287
           :::          |||    |||:::    |||    :::         :::
    4694 ATGAAGATATA...GGTGAAGACAAAGATGAAGTTATAGATTTAATAGTC 4740

     288 GluAspGluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAs 304
         :::       |||:::          ::::::::||||||    ::::::|||::
    4741 CAAAAAGAGAAACGCATTGAAAAGGTTAAAGCGAAAAAGAAAAAATTAGA 4790

     304 pGlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuV 321
         ::::     |||:::           |||    ::::::    |||    :::    |
    4791 AAAAAAAGTTGAA.........GAAGGTGTTAGTGGTCTTAAAAAACACG 4831

     321 alLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsn 337
          ||    :::      ||||||    |||        ::::|||    :::||||||:::
    4832 TAGACGAAGTAATGAAATATGTTCAAAAAATTGATAAAGAAGTTGATAAA 4881

     338 GlyLysThrLysAspLeu......TyrArgAspLeuIleAsnIlePheAs 352
                :::|||    |||         :::|||:::    |||::::::
    4882 GAAGTATCTAAAGCTTTAGAATCAAAAAATGAIGTTACTAATGTTTTAAA 4931

     352 pLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetP 369
          :::                                   :::    :::|||||
    4932 ACAA........................AATCAAGATTTTTTT.... 4950

     369 heTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHis 385
                                          :::    ::::::::
    4951 ..................................AGTAAAGTTAAAAAC 4965

     385 LeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAl 402
         :::    ||||||    :::          :::    |||    |||         ||
    4966 TTCGTAAAAAAATATAAAGTATTTGCTGCACCATTCATATCTGCCGTTGC 5015

     402 aAla...GlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleP 418
         ||||     ::::::|||:::    :::|||::::::    :::    |||
    5016 AGCATTTGCATCATATGTAGTTGGTTCTTTACATTTTCTTTATTTTCAT 5065

     418 roLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHis 434
              :::|||:::    |||         ||||||    :::
    5066 CATGTGTAACAATAGCTTCTTCAACTTACTTATTATCAAAAGTTGACAAA 5115

     435 IleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAl 451
                :::|||:::::
    5116 ACTATAAATAAAAATAAGGAGAGA......................... 5139

     451 aLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheV 468
                 |||||||||||                          ||||
    5140 .........CCGTTTTATTCA........................TTTG 5155

     468 alPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGln 484
          |||||          :::          |||||||||           :::
    5156 TATTTGAT............ATCTTTAAGAATTTAAAA.........CAT 5184

     485 TyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnPr 501
         |||||||||::::::::    :::                ::::::    |||
    5185 TATTTACAACAAATGAAAGAAAAATTTAGTAAAGAAAAAAATAATAATGT 5234

     501 oLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrA 518
          ::::::::          :::          :::       :::    ::::|||
    5235 AATAGAAGTAACAAACAAAGCTGAGAAAAAAGGTAATGTACAGGTAACAA 5284

     518 snLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArg 534
          |||||     :::    :::          |||     ::::::::|||||||:::
    5285 ATAAAACCAGAAAACA.........ACTAAAGTTGATAAAAATAATAAA 5325

     535 GlnMetLeuProValIleArgSerThr 543
            ::::|||     |||   |||
    5326 . ....GTACCGAAAAAAAGAAGAACG 5346

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-973-462 1
```

```
seq documentation block:
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
;  APPLICANT: DRUILHE, PIERRE
;  APPLICANT: DAUBERSIES, PIERRE
;  TITLE OF INVENTION: MALARIAL PRE ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
;  FILE REFERENCE: 0660 0125 0 PCT
;  CURRENT APPLICATION NUMBER: US/08/973,462B
;  CURRENT FILING DATE: 1998 02 06
;  EARLIER APPLICATION NUMBER: PCT/FR96/00894
;  EARLIER FILING DATE: 1996 06 12
;  EARLIER APPLICATION NUMBER: FR 95/07007
;  EARLIER FILING DATE: 1995 06 13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 6152
;   TYPE: DNA
;   ORGANISM: P. falciparum
US 08 973 462 1

alignment scores:
            Quality:    99.50            Length:     559
              Ratio:   0.381               Gaps:      23
 Percent Similarity:  46.691   Percent Identity:  19.499

alignment block:
 US 09 932 678 2 x US 08-973 462-1    ..

 Align seq 1/1  to: US 08-973 462 1  from: 1  to: 6152

    31 IleSerAsnMetArgAlaLeuGluAsnAspPhePheAsnSerProProAr 47
       :::|||::::::::::    ||||||:::|||:::::::::::      ::
  4193 GTATCTGATTTAAAAGATCTTGAAGAAGATATATTAAAAGAAGTAAAAGA 4242

    47 g.........................................LysT 49
       :                                          ||||
  4243 AATCAAAGAACTTGAAAGTGAAATTTTAGAAGATTATAAAGAATTAAAAA 4292

    49 hrValArgPheGlyGlyThrValThrGluValIleuLeuLysTyrLysLys 65
       ||::::::                   |||:::::::|||    :::   ||||:::
  4293 CTATTGAA...............ACAGATATTTTAGAAGAGAAAAAAGAA 4327

    66 GlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAs 82
          |||     :::     ||||||    ::::::::::::         ::
  4328 ATAGAAAAAGATCATTTTGAAAAATTCCAAGAAGAAGCT......GAAGA 4371

    82 pIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSer.... 97
       :|||||||||     :::     :::   |||   |||     |||
  4372 AATAAAAGATCTTGAA...GCAGATATATTAAAAGAAGTATCTTCATTAG 4418

    98 ..........................SerIleMetTyrLeuThrLysAsp 105
                                 ::::::     |||   ::::::
  4419 AAGTTGAAGAAGAAAAAAAATTACAAGAAGTACACGAATTAAAAGAAGAG 4468

   106 PheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgSe 122
          |||::::::||||||          ::::::      |||:::::::::
  4469 GTAGAACATATAATAAGTGGTGATGCCATATAAAAGCTTTGGAAGAAGA 4518

   122 rGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerA 139
       ::::        :::||||||          :::    |||::::::
  4519 TGAT......TTAGAAGAAGTAGATGATTTAAAAGGAAGTATA....... 4555

   139 laGlnThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPhe 155
                                 |||:::|||:::
  4556 ..........................TTAGACATGTTA............ 4567

   156 ValProProArgValIleIleLysGluGlyAspValAspValSerAspSe 172
                               :::||||||:::::::::::::::|||
  4568 ........................AAGGCAGATATGGAATTAGGGGATAT 4593

   172 rAspAspGluAsp.........................AspAsnLeuP 180
        |||   |||:::                          :::::::::
  4594 GGATAAGGAAAGTTTAGAAGATGTAACAACAAAACTTGGAGAAAGAGTTG 4643

   180 roAlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyr 196
         :::         |||:::          ||||||     :::      :::
  4644 AATCCTTAAAAGATGTTTTATCTAGTGCATTAGGCATGGATGAAGAACAA 4693

   197 ValProSerThrProTrpPheLeuMetProIleLeuValGluLysPhePr 213
       :::   :::                   |||   |||   |||   :::
  4694 ATGAAAACAAGAAAAAAAGCTCAAAGACCTAAGTTGGAAGAAGTATTATT 4743

   213 oPheValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeuL 230
              :::   :::|||:::       :::      :::   :::   :
  4744 AAAAGAAGAGGTTAAAGAAGAACCAAAGAAAAAAATAACAAAAAAGAAAG 4793

   230 euArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLeu 246
        ::|||::::::::::                 :::   ||||||:::|||:::
  4794 TAAGGTTTGATATTAAGGATAAGGAACCAAAAGATGAAATAGTACAAGTT 4843

   247 ...........................IleIleGluLysLeuLeuLysLe 254
                                     :::|||:::   ::::::
  4844 GAAATGAAAGATGAAGATATAGAAGAAGATGTAGAAGAAGATATAGAAGA 4893

   254 uAspValAsnAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaT 271
        |||::::::   ::::::         :::||||||    :::|||
  4894 AGATATAGAAGAAGATAAA......GTTGAAGATATAGATGAAGATATAG 4937

   271 hrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAsp 287
         :::             |||   |||:::   |||     :::      :::
  4938 ATGAAGATATA...GGTGAAGACAAAGATGAAGTTATAGATTTAATAGTC 4984

   288 GluAspGluGluThrGluHisGluThrLysAlaGlyProGluArgLeuAs 304
       :::   |||:::         :::::::::||||||        ::::::|||::
  4985 CAAAAAGAGAAACGCATTGAAAAGGTTAAAGCGAAAAAGAAAAAATTAGA 5034

   304 pGlnMetValHisProValAlaGlnArgLeuAspIleLeuMetSerLeuV 321
       ::::    |||:::           |||   ::::::   |||   :::   |
  5035 AAAAAAAGTTGAA.........GAAGGTGTTAGTGGTCTTAAAAAACACG 5075

   321 alLeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsn 337
       |:   :::   ||||||   |||          :::|||   :::|||||||:::
  5076 TAGACGAAGTAATCAAATATGTTCAAAAAATTGATAAAGAAGTTGATAAA 5125

   338 GlyLysThrLysAspLeu.....TyrArgAspLeuIleAsnIlePheAs 352
             :::|||   |||            :::|||:::   |||::::::
  5126 GAAGTATCTAAAGCTTTAGAATCAAAAAATGATGTTACTAATGTTTTAAA 5175

   352 pLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMetP 369
        :::                             :::   :::||||||
  5176 ACAA..........................AATCAAGATTTTTT.... 5194

   369 heTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluHis 385
                                           :::   :::::::::
  5195 ................................AGTAAAGTTAAAAAC 5209

   386 LeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGlnAl 402
       :::   ||||||    :::        :::   |||   |||          ||
  5210 TTCGTAAAAAAATATAAAGTATTTGCTGCACCATTCATATCTGCCGTTGC 5259

   402 aAla...GlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIleP 418
       ||||     ::::::|||:::     :::|||::::::      :::    |||
  5260 AGCATTTGCATCATATGTAGTTGGGTTCTTTACATTTTCTTTATTTTCAT 5309

   418 roLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHis 434
            :::|||:::    |||         ||||||    :::
  5310 CATGTGTAACAATAGCTTCTTCAACTTACTTATTATCAAAAGTTGACAAA 5359

   435 IleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAl 451
               :::|||::::::
  5360 ACTATAAATAAAAATAAGGAGAGA........................ 5383

   451 aLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheV 468
```

```
                 ||||||||||||                          ||||
    5384 .........CCGTTTTATTCA.......................TTTG 5399

     468 alPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGln 484
         |||||                 :::      |||||||||         :::
    5400 TATTTGAT............ATCTTTAAGAATTTAAAA.........CAT 5428

     485 TyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnPr 501
         |||||||||:::::::::   :::              ::::::   |||
    5429 TATTTACAACAAATGAAAGAAAAATTTAGTAAAGAAAAAAATAATAATGT 5478

     501 oLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrA 518
          :::::::::          :::      :::      :::   :::||||
    5479 AATAGAAGTAACAAACAAAGCTGAGAAAAAAGGTAATGTACAGGTAACAA 5528

     518 snLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArg 534
         |||||   :::   :::          |||   ::::::::||||||:::
    5529 ATAAAACCGAGAAAACA.........ACTAAAGTTGATAAAAATAATAAA 5569

     535 GlnMetLeuProValIleArgSerThr 543
             :::|||      |||   |||
    5570 ......GTACCGAAAAAAAGAAGAACG 5590

seq_name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-776-265-4

seq_documentation_block:
; Sequence 4, Application US/08776265
; Patent No. 6001631
;  GENERAL INFORMATION:
;    APPLICANT:  BLANCHE, Francis
;    APPLICANT:  CAMERON, Beatrice
;    APPLICANT:  CROUZET, Joel
;    APPLICANT:  FAMECHON, Alain
;    APPLICANT:  FERRERO, Lucia
;    TITLE OF INVENTION:  No. 6001631el Topoisomerase IV, Corresponding
;    TITLE OF INVENTION:  Nucleotide Sequences and Uses Thereof
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner, L.L.P.
;      STREET:  1300 I. Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/776,265
;      FILING DATE:  24-JAN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Einaudi, Carol P.
;      REGISTRATION NUMBER:  32,220
;      REFERENCE/DOCKET NUMBER:  03806.0394-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4444
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2402 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-776-265-4


alignment_scores:
           Quality:   98.50            Length:    470
           Ratio:  0.462               Gaps:     19
 Percent Similarity:  45.319   Percent Identity: 17.660

alignment_block:
 US-09-932-678-2 x US-08-776-265-4  ..

 Align seq 1/1  to: US-08-776-265-4  from: 1  to: 2402

      79 LeuAspProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPh 95
         |||       |||:::::::::::::::::|||   |||:::|||      ::
     906 TTGAAAAAAGATGTGAACAGTGAATCAATCAAAAATTATCTT......TA 949

      95 eArgSerSerIleMetTyrLeuThrLysAspPheGluGlnLeuIleSerI 112
         ::::::::|||   :::   ::::::  :::|||  :::::::::::::|
     950 TAAAAACTCTGATTTACAGATTTCATATAATTTC...AACATGGTCGCTA 996

     112 IeIleLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGlu 128
         ||          |||   |||         :::  ::::::::::::
     997 TTAGTGATGGTCGTCCAAAATTGATGGGTATTCGTCAAATTATAGATAGT 1046

     129 TyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuAr 145
         ||||||            ::::::|||:::   :::|||   |||
    1047 TATTTGAATCATCAAATTGAGGTTGTTGCAAATAGAACGAAGTTT..... 1091

     145 qProCysLeuSerMetIleAlaSerHisPheValProProArqValIleI 162

    1091 . .. ..................................................... 1091

     162 IeLysGluGlyAspValAspValSerAspSerAspAspGluAspAspAsn 178

    1091 ..... 1091

     179 LeuProAlaAsnPheAspThrCysHisArqAlaLeuGlnIleIleAlaAr 195
              :::::::|||:::::::::::::   :::::::|||:::
    1092 .........GAATTAGATAATGCTGAAAAACGTATGCATATCGTTGAAGG 1132

     195 qIyrValProSerThrProTrpPheLeuMetProIleLeuValGluLysP 212
             :::   :::                :::   ||||||   :::   :
    1133 TTTGATTAAAGCG.............TTGTCAATTTTACATAAAGTAA 1167

     212 hsProPheValArqLysSerGluArqThrLeuGluCysTyrValHisAsn 228
         ::   ::::::|||:::|||:::::::      ::::::      ::||||
    1168 TCGAATTGATTCGTAGCTCTAAAAACAAGCGTGACGCT...AAAGAAAAC 1214

     229 LeuLeuArqIleSerValTyrPheProThrLeuArqHisGluIleLeuGl 245
         |||:::::::::      :::                 :::   |||:::
    1215 CTTATCGAAGTATACGAGTTCACAGAAGAACAGGCTGAAGCAATTGTAAT 1264

     245 uLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArqGlnG 262
         :||   :::   :::|||   :::   |||:::   |||   :::
    1265 GTTACACTTATATCGTTTAACAAACACTGACATAGTTGCGCTTGAAGGTG 1314

     262 lyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAsp 278
         ::::::   |||              |||                 |||
    1315 AACATAAAGAACTTGAAGCATTAATCAAACAATTACGTCATATTCTTGAT 1364

     279 SerThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisCl 295
         :::   ::::::|||:::|||:::   :::::::|||   :::|||   ::
    1365 AACCATGATGCATTATTGAATGTCATAAAAGAAGAATTGAATGAAATTAA 1414

     295 uThrLysAlaGlyProGluArqLeuAspGlnMetValHisProValAlaG 312
         :  |||      ||||:|||||:::  :::      :::   |
    1415 AAAGAAATTCAAATCTGAACGACTGTCTTTAATTGAAGCAGAAATTGAAG 1464

     312 lu...ArqLeuAsp.........................IleLeuMet 328
         ||   :::::|||                          :::::::::
    1465 AAATTAAAATTGACAAAGAAGTTATGGTGCCTAGTGAAGAAGTTATTTTA 1514

     329 SerLeuVal.......................................... 321
         |||::::::
    1515 AGTATGACACGTCATGGATATATTAAACGTACTTCTATTCGTAGCTTTAA 1564

     322 .LeuSerTyrMetLysAspValCysTyrValAspGly............L 334
```

```
          |||   ::::::|||::            ||||||          |
1565 TGCTAGCGGTGTTGAAGATATTGCTTTAAAAGATGGTGACAGTTTACTTA 1614

 334 ysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIle 350
     ||    ::::::   :::|||::|||              :::   |||
1615 AACATCAAGAACTAAATACCCAAGATACCGTACTAGTATTTACAAATAAA 1664

 351 PheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePh 367
              |||::::::|||:::|||            ::::::
1665 GGTCGTTATCTATTTATACCAGTTCATAAATTACGAGATATTCGT..... 1709

 367 eMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuG 384

1709 .................................................. 1709

 384 luHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArg 400
            ||||||:::|||
1710 .......TCGAAAGAATTG............................... 1721

 401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIl 417
            |||::::::::                     ::::::::::
1722 .........GGGCAACATGTA..................TCACAAATAGT 1744

 417 eProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuH 434
     :|||:::                      |||    :::|||        ::::
1745 TCCTATCGAAGAA...............GATGAAGTGGTT......ATTA 1773

 434 isIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspVal 450
     ::::::|||    |||:::::::|||    :::|||    ||||||
1774 ATGTTTAT...AATGAAAAGGACTTTAATACTGATGCATTTATGTTTTT 1820

 451 AlaLeu...HisGlyProPheTyrSerAlaCysGlnAlaValPheTyrTh 466
     |||        :::|||    :::    ::::::         :::|||    ||
1821 GCGACTCAAAATGGCATGATTAAGAAAAGTATAGTGCCTCTATTTAAAAC 1870

 466 rPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyL 483
     |         |||      :::|||    |||::::::    ::::::|||||||:::
1871 AACGCGTTTT...AATAAACCTTAATTGCAACTAAAGTTAAAGAAAATG 1917

 483 euGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeu 499
       :::    ::::::    ::::::||||||:::            :::||||||
1918 ATGATTGATTACTGTTATGCGTTTGAAAAA.........GATCAATTA 1958

 500 AsnProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIl 516
                                            :::::::::||
1959 .........................................ATTACCGTAAT 1969

 516 eThrAsnLys 519
     ||||||||||
1970 TACAAATAAA 1979
```

seq name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-776-265-1

```
seq documentation block:
; Sequence 1, Application US/08776265
; Patent No. 6001631
;  GENERAL INFORMATION:
;     APPLICANT:  BLANCHE, Francis
;     APPLICANT:  CAMERON, Beatrice
;     APPLICANT:  CROUZET, Joel
;     APPLICANT:  FAMECHON, Alain
;     APPLICANT:  FERRERO, Lucia
;     TITLE OF INVENTION:  No. 6001631el Topoisomerase IV, Corresponding
;     TITLE OF INVENTION:  Nucleotide Sequences and Uses Thereof
;     NUMBER OF SEQUENCES:  17
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;       ADDRESSEE:  Dunner, L.L.P.
;       STREET:  1300 I. Street, N.W., Suite 700
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  USA
;       ZIP:  20005-3315
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/776,265
;       FILING DATE:  24-JAN-1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Einaudi, Carol P.
;       REGISTRATION NUMBER:  32,220
;       REFERENCE/DOCKET NUMBER:  03806.0394-00000
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 408-4000
;       TELEFAX:  (202) 408-4444
;  INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4565 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-776-265-1
```

```
alignment_scores:
            Quality:    98.50              Length:     470
              Ratio:    0.462                Gaps:      19
 Percent Similarity:   45.319    Percent Identity:  17.660

alignment_block:
 US-09-932-678-2 x US-08-776-265-1

 Align seq 1/1  to: US-08-776-265-1  from: 1  to: 4565

   75 LeuAspProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPh 95
         |||       |||::::::::::::::::|||    |||:::|||        ::
 2933 TTGAAAAAAGATGTGAACAGTGAATCAATCAAAAATTATCTT......TA 2981

   95 eArgSerSerIleMetTyrLeuThrLysAspPheGluGlnLeuIleSerI 112
      ::::::::|||    :::    ::::::    :::|||    ::::::::::::|
 2982 TAAAAACTCTGATTTACAGATTTCATATAATTTC...AACATGGTCGCTA 3028

  112 leIleLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGlu 128
      ||               |||    |||           :::    :::::::::::
 3029 TTAGTGATGGTCGTCCAAAATTGATGGGTATTCGTCAAATTATAGATAGT 3078

  129 TyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuAr 145
      ||||||                ::::::|||:::    :::|||    |||
 3079 TATTTGAATCATCAAATTGAGGTTGTTGCAAATAGAACGAAGTTT..... 3123

  145 gProCysLeuSerMetIleAlaSerHisPheValProProArgValIleI 162

 3123 .................................................. 3123

  162 leLysGluGlyAspValAspValSerAspSerAspAspGluAspAspAsn 178

 3123 .................................................. 3123

  178 LeuProAlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaAr 195
                :::::::|||::::::::::::    ::::::::|||:::
 3124 .........GAATTAGATAATGCTGAAAAACGTATGCATATCGTTGAAGG 3164

  195 gTyrValProSerThrProTrpPheLeuMetProIleLeuValGluLysP 212
           :::     :::                 :::   ||||||    :::     :
 3165 TTTGATTAAAGCG..............TTGTCAATTTTAGATAAAGTAA 3199

  212 heProPheValArgLysSerGluArgThrLeuGluCysTyrValHisAsn 228
       :.    ::::::|||:::|||::::::            ::::::         :::|||
 3200 T.GAATTGATTCGTAGCTCTAAAAACAAGCGTGACGCT...AAAGAAAAC 3246

  229 LeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGl 245
```

```
      |||:::::::::        :::                   :::    |||:::
3247 CTTATCGAAGTATACGAGTTCACAGAAGAACAGGCTGAAGCCAATTGTAAT 3296

 245 uLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnG 262
      |||    :::    :::|||    :::    |||:::    |||    :::
3297 GTTACAGTTATATCGTTTAACAAACACTGACATAGTTGCGCTTGAAGGTG 3346

 262 lyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAsp 278
      ::::::    |||              |||                |||
3347 AACATAAAGAACTTGAAGCATTAATCAAACAATTACGTCATATTCTTGAT 3396

 279 SerThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGl 295
      :::    ::::::|||:::|||:::    ::::::|||    ::::|||    ::
3397 AACCATGATGCATTATTGAATGTCATAAAAGAAGAATTGAATGAAATTAA 3446

 295 uThrLysAlaGlyProGluArgLeuAspGlnMetValHisProValAlaG 312
      :    |||           |||||||||:::    :::         :::    |
3447 AAAGAAATTCAAATCTGAACGACTGTCTTTAATTGAAGCAGAAATTGAAG 3496

 312 lu...ArgLeuAsp..........................IleLeuMet 318
      ||    ::::::|||                          :::::::::
3497 AAATTAAAATTGACAAAGAAGTTATGGTGCCTAGTGAAGAAGTTATTTTA 3546

 319 SerLeuVal......................................... 321
      |||::::::
3547 AGTATGACACGTCATGGATATATTAAACGTACTTCTATTCGTAGCTTTAA 3596

 322 .LeuSerTyrMetLysAspValCysTyrValAspGly............L 334
         |||    ::::::|||:::        ||||||                |
3597 TGCTAGCGGTGTTGAAGATATTGGTTTAAAAGATGGTGACAGTTTACTTA 3646

 334 ysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIle 350
      ||    ::::::    :::|||:::|||                :::    |||
3647 AACATCAAGAAGTAAATACGCAAGATACCGTACTAGTATTTACAAATAAA 3696

 351 PheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePh 367
                |||::::::|||:::|||               ::::::
3697 GGTCGTTATCTATTTATACCAGTTCATAAATTACGAGATATTCGT..... 3741

 367 eMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuG 384

3741 ................................................. 3741

 384 luHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArg 400
                |||||||:::|||
3742 ........TGGAAAGAATTG............................. 3753

 401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIl 417
                |||::::::::                    ::::::::::
3754 .........GGGCAACATGTA..................TCACAAATAGT 3776

 417 eProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuH 434
      :|||:::                          |||    :::|||      ::::
3777 TCCTATCGAAGAA..............GATGAAGTGGTT......ATTA 3805

 434 isIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspVal 450
      ::::::|||    |||:::::::|||    :::|||    ||||||
3806 ATGTCTAT...AATGAAAAGGACTTTAATACTGATGCATTTTATGTTTTT 3852

 451 AlaLeu...HisGlyProPheTyrSerAlaCysGlnAlaValPheTyrTh 466
      |||          :::|||    :::    ::::::             :::|||    ||
3853 GCGACTCAAAATGGCATGATTAAGAAAAGTACAGTGCCTCTATTTAAAAC 3902

 466 rPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyL 483
      |        |||    :::|||    |||:::::::    ::::::|||||||:::
3903 AACGCGTTTT...AATAAACCTTTAATTGCAACTAAAGTTAAAGAAAATG 3949

 483 euGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeu 499
        :::    ::::::    ::::::||||||:::            :::||||||
3950 ATGATTTGATTAGTGTTATGCGCTTTGAAAAA.........GATCAATTA 3990

 500 AsnProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIl 516
                                             :::::::::||
```

```
3991 ........................ ...............ATTACCGTAAT 4001

 516 eThrAsnLys 519
      |||||||||
4002 TACAAATAAA 4011
```

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US 08 417 330A 13

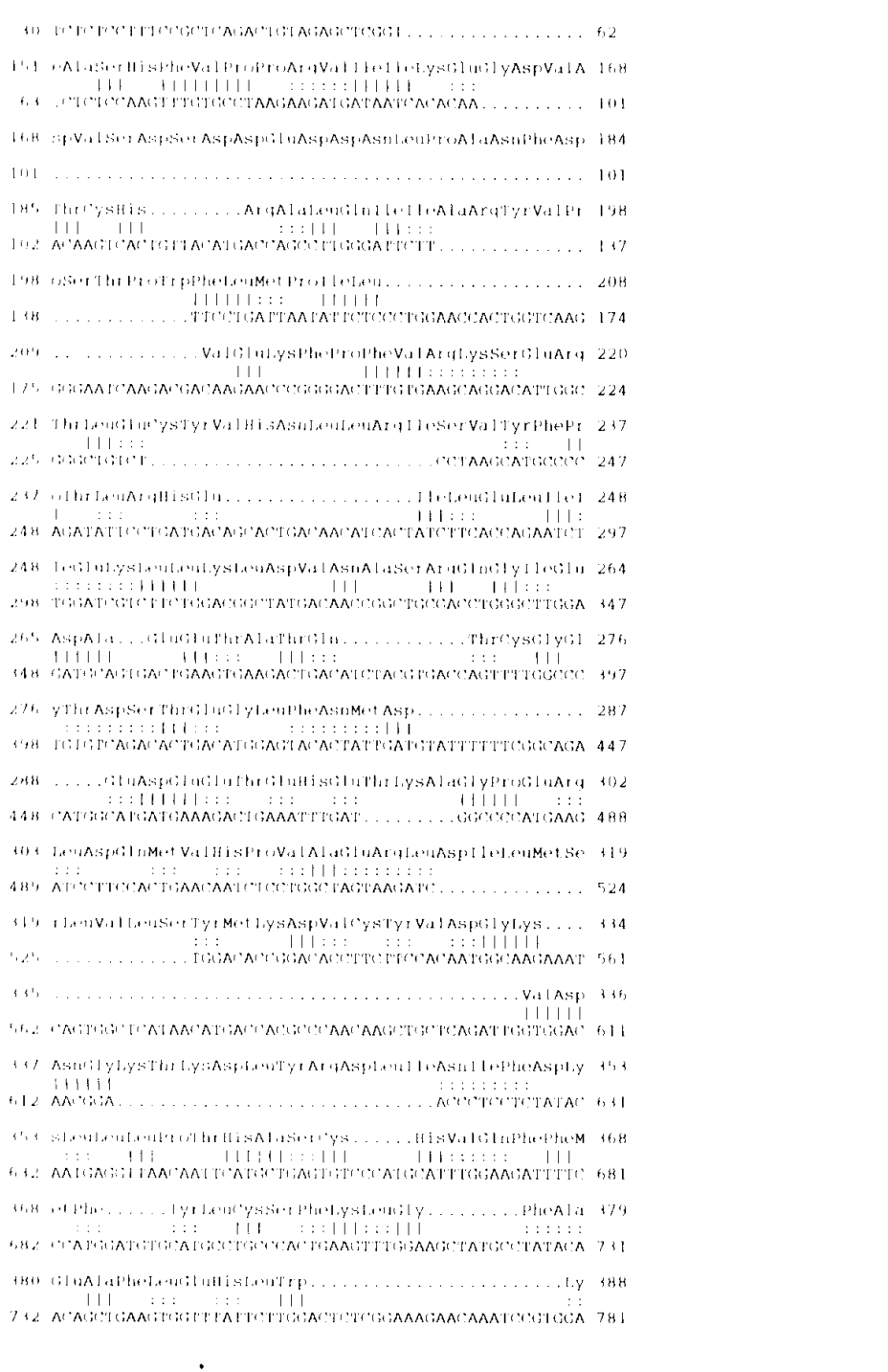
seq_documentation_block:
; Sequence 13, Application US/08417330A
; Patent No. 5719057
; GENERAL INFORMATION:
; APPLICANT: HADINGHAM, KAREN
; APPLICANT: LE BOURDELLES, BEATRICE
; APPLICANT: WHITING, PAUL
; APPLICANT: WINGROVE, PETER
; TITLE OF INVENTION: STABLY TRANSFECTED CELL LINE EXPRESSING
; TITLE OF INVENTION: GABA-A RECEPTOR AND NOVEL CLONED
; TITLE OF INVENTION: GABA-RECEPTOR SUBUNIT cDNA SEQUENCES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: J. MARK HAND - MERCK & CO., INC.
; STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
; CITY: RAHWAY
; STATE: NJ
; COUNTRY: US
; ZIP: 07065-0900
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417,330A
; FILING DATE: 05-APR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: HAND, MARK
; REGISTRATION NUMBER: 36,545
; REFERENCE/DOCKET NUMBER: T1109IA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-594-3905
; TELEFAX: 908-594-4720
; TELEX:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1638 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 87...1562
; OTHER INFORMATION:
US-08-417-330A-13

alignment_scores:
Quality: 97.00 Length: 626
Ratio: 0.406 Gaps: 31
Percent Similarity: 38.179 Percent Identity: 18.051

alignment_block:
US-09-932-678-2 x US 08-417-330A 13 ..

Align seq 1/1 to: US-08-417-330A 13 from: 1 to: 1638

```
 135 AsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMetIl 151
        :::|||:::||||.|||||||||    |||
```



```
     641 oProValLeuTyrMetGlnProSerPro 650
         |   :::   |||:::|||   ||||||
    1388 CAAGGCCACCTACGTGCAGGACAGCCCG 1415

seq_name: /cqn2_6/ptodata/1/ina/5B_COMB.seq:US-08-680-327-2

seq_documentation_block:
; Sequence 2, Application US/08680327
; Patent No. 5859321
;  GENERAL INFORMATION:
;     APPLICANT:  Staskawicz, Brian S., Oldroyd, Giles Edward,
;     APPLICANT:  Salmeron, John M., Rommens, Caius
;     TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR PLANT
;     TITLE OF INVENTION:  PATHOGEN RESISTANCE
;     NUMBER OF SEQUENCES:  15
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Klarquist Sparkman Campbell Leigh &
;       ADDRESSEE:  Whinston
;       STREET:  One World Trade Center
;       STREET:  121 S.W. Salmon Street
;       STREET:  Suite 1600
;       CITY:  Portland
;       STATE:  Oregon
;       COUNTRY:  United States of America
;       ZIP:  97204
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Disk, 3-1/2 inch
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  MS DOS
;       SOFTWARE:  WordPerfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/680,327
;       FILING DATE:  July 11, 1996
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/310,912
;       FILING DATE:  September 22, 1994
;       CLASSIFICATION:  800
;       APPLICATION NUMBER:  08/227,360
;       FILING DATE:  April 13, 1994
;       CLASSIFICATION:  800
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Dow, Alan. E.
;       REGISTRATION NUMBER:  35,123
;       REFERENCE/DOCKET NUMBER:  5151-45038
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (503) 226-7391
;       TELEFAX:  (503) 228-9446
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10968 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double stranded
;       TOPOLOGY:  linear
US-08-680-327-2


alignment_scores:
            Quality:   97.00            Length:    699
              Ratio:  0.344              Gaps:     29
 Percent Similarity:  40.343   Percent Identity: 16.452

alignment_block:
 US-09-932-678-2 x US-08-680-327-2   ..

 Align seg 1/1  to: US-08-680-327-2  from: 1  to: 10968

      102 LeuThrLysAspPheGluGlnLeuIleSerIle...IleLeuArgLeuPr 117
          :::   |||   |||::::::   |||   :::      ||||||
     5484 ATATTGAAGAGGTTTCATGAATATATTCTTGTCAATCGTCTACGGAAGGA 5533

      117 oTrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLeuG 134
                |||   :::   |||::::::::::|||         |||

5534 TGAAACCAATTTGACATTTACTATTGCAGATGAGGTCAAAAAGTTTTATC 5583

 134 lyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMet 150
       :::|||:::      |||   :::::::::|||
5584 ATGGGTTGTTGCTCATGGTTACATATCTTATTGAACCTCCA..... 5624

 151 IleAlaSerHisPheValProProArgValIleIleLysGluGlyAspVa 167
                  ||||||              |||:::::|||::
5625 .............GTTCCTCACACTGAATGCAGGAAGCAAAATGATCT 5669

 167 lAspValSerAspSerAsp..................................... 173
     :::::::      :::
5660 CTCAATGCGACATGAAGCTGTTGCAATTGAGGCGGAATCTGCTGTGTGTT 5709

 174 ..AspGluAspAspAsnLeuProAlaAsnPheAspThrCysHisArgAla 189
           :::||||||||::      |||              |||
5710 TACATTATGAGGATAATATGAATAACAAC.....AGTAGGGAG 5747

 190 LeuGlnIleIleAlaArgTyrValProSerThrProTrpPhe..... 203
     ::::::   :::   :::::::::      |||   ||::::
5748 ATCAATCAGGTACTTCAGTTTTTGACTGTGACTTTCTGGCTTATCAAGTC 5797

 204 .........LeuMetProIleLeuValGluLysPheProPheValArgL 217
             ||||||   :::|||   :::|||      :::   ::::
5798 TGACGCTAACTTGATGGATCTACTGAAGCACAAATCCACTTTGGGAAATC 5847

 217 ysSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSer 233
     ::          :::|||         ::::::|||:::
5848 AAGTTCTAGATCTGATTGAGAGTGCTCATGAAGAGCTTATT..... 5888

 234 ValTyrPheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLy 250
                  |||||      :::::::::|||:::   :::||
5889 ...........CTGCTTAGATCTATTCTCATGGATCTTCTTAGGAAAAA 5926

 250 sLeuLeuLysLeuAsp.....................ValAsnAlaSerA 260
     |||   :::||||||                      |||:::|||:::|
5927 GCTTTACAGATTGGATGATCTCTTAATGCATGCTGAGGTGACTGCAAAAA 5976

 260 rq..................................................... 260
     ||
5977 GGTTAGCAATATTCAGTGGTTCTTGTTATGAATATTTCATGAACGGAAGC 6026

 261 .....................................GlnGlyIleGl 264
                                          |||  |||||
6027 AGCACTGAGAAAATGAGGCCCTTGTTATCTGATTTCTGCAAGAGATTGA 6076

 264 uAspAlaGluGluThrAlaThrGlnThrCys..................... 274
     |:::::::::          ::::::|||
6077 GTCTGTCAAGGTAGAGTTCAGAAATGTTTGCTTGCAAGTTCTGGATATAT 6126

 275 ........GlyGlyThrAspSerThrGluGlyLeuPheAsn......... 285
             :::   ||||||:::   ||||||||||   |||
6127 CACCTTTTTCCCTGACAGATGGA...GAAGGCCTTGTTAATTTCTTATTA 6173

 285 ..................................................... 285

6174 AAAAACCAGGCCAAGGTGCCGAATGATGATGCTGTTTCTTCTGATGGAAG 6223

 286 .MetAspGluAspGluGluThrGlnHis....................... 294
     :::::::::   ::::::::::|||
6224 TTTAGAGGATGCAAGCAGCACTGAGAAAATGGGACTTCCATCTGATTTTC 6273

 295 ....................GluThrLysAlaGlyProGluArgLeuAsp 304
                         |||   |||   :::   :::      |||
6274 TCCGAGAGATTGAGTCTGTTGAGATAAAGGAGGCCAGAAAATTATATGAT 6323

 305 GlnMetVal.........HisProValAlaGluArgLeuAsp......Il 316
     |||::::::         |||      :::::::::   |||
6324 CAAGTTTTGGATGCAACACATTGTGAGACGAGTAAGACAGATGCAAAAAG 6373

 316 eLeuMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspG 333
     ::::::::::::::::|||:::   :::      :::
6374 CTTTATCAACATTATGTTAACCCAACAGGACAAGTTG.....C 6411
```

```
 333 lyLysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsn 349
              |||    |||::::::::    |||     :::::::::
6412 CGGACTATGATGCTGGTTCACTCTCTTATCTTCTTAACCAAATATCAGTA 6461

 350 IlePheAspLysLeuLeu.......................LeuProTh 358
     :::    ||||||||||||                       :::
6462 GTTAAAGACAAACTATTGCACATTGCCTCTTTACTTGTAGATATTGTACA 6511

 358 rHisAlaSerCysHisValGln........................... 365
      :::    :::    |||::::::
6512 CTACCGGAATATGCATATAGAACTTACACATCTCGCTGAACGTGTTCAAG 6561

 365 ................................................. 365

6562 ATAAAAACTACATTTGTTTCTTCTCTGTCAAGGGTTATATTCCTGCTTGG 6611

 366 PhePheMetPheTyrLeuCysSerPheLys......LeuGlyPheAlaGl 380
     ::::::    :::||||||    :::    |||       |||   |||:::||
6612 TATTACACACTATATCTCTCTGATGTCAAGCAATTGCTTAAGTTTGTTGA 6661

 380 nAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSerAsnProA 397
     ||||     ::::::    :::       |||:::    |||    |||:::    :
6662 GGCAGAGGTAAAGATTATTTGTCTGAAAGTACCAGATTCTTCAAGTTATA 6711

 397 laIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArg 413
     :::::    :::       :::    :::|||::::::     ||||||::::::
6712 GCTTCCCTAAGACAAATGGATTAGGATATCTCAATTGCTTTTTAGGCAAA 6761

 414 AlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeuVa 430
        :::    :::                :::::::|||    |||||||||||:::::
6762 TTCGAGGAGCTT...........TTACGTTCTAAGCTCGATTTGATAAT 6799

 430 lAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaP 447
     ::::       |||
6800 CGACTTAAAACAT.................................... 6812

 447 heCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaVal 463

6812 ................................................. 6812

 464 PheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLy 480
                                  |||:::    |||        :::||
6813 ..........................CAGATTGAATCA......GTCAA 6829

 480 sGluGlyLeuGlnTyrLeuGlnSerLeu......AsnPheGluArgIle. 494
     ||||||||||        |||:::|||:::          :::|||::::::
6830 GGAGGCTTATTGTGCCTAAGATCATTCATTGATCATTTTTCAGAAAGCT 6879

 495 ..........................ValMetSerGlnLeuAsnProLeu 502
                               ::::::::::::::::::::    :::
6880 ATGATGAGCATGATGAAGCTTGTGGTCTTATAGCAAGAGTTTCTGTAATG 6929

 503 LysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAsnLy 519
                        |||::::::            |||     ::::::
6930 GCATACAAGGCTGAGTATGTCATTGACTCATGCTTGGCCTATTCTCATCC 6979

 519 s......TyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnA 534
             |||::::::::::::::          :::::::|||     |||     :
6980 ACTCTGGTACAAAGTTCTTTGGATTTCTGAACTTCTTGAG...AATATTA 7026

 534 rgGluMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGln 550
     ::     :::          |||:::      :::|||:::              :::::::::
7027 AGCTTGTAAATAAAGTTGTTGGTGACACATGTGAAAGAAGGAACATTGAA 7076

 551 IleCysThrAsnProLeuAspThrPhePheProPheAspProCysValIle 567
     :::    ::::::                                          ::
7077 GTTACTGTGCATGAA.............................GT 7093

 567 uLysArgSerLysLysPheIleAspProIleTyrGlnValTrpGluAspM 584
     :    ::::::         ::::::    |||     ::::::::::::    :::
7094 TGCAAAGACTACCACTTATGTAGCACCATCTTTTTCAGCTTATACTCAAA 7143

 584 etSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle 600
         :::    |||:|||::::::    |||:::         :::
7144 GAGCAAACGAAGAAATGGAGGGTTTTCAGGATACAATA............ 7181

 601 ValGluAspGluAspAspAspPheLeuLysGlyGluValProGlnAsnAs 617
           ||||||        |||    |||    |||        |||:::    ||
7182 ......GATGAATTAAAGGATAAACTACTTGGAGGATCACCTGAGCTTGA 7225

 617 p......ThrValIleGlyIle............................ 622
     |         ::::::::::|||:::
7226 TGTCATCTCAATCGTTGGCATGCCAGGATTGGGCAAGACTACACTAGCAA 7275

 623 .........................ThrProSerSerPheAspThrHis 630
                              :::    |||    ||||||:::|||
7276 AGAAGATTTACAATGATCCAGAAGTCACCTCTCGCTTCGATGTCCAT 7322
```

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-082-649B-80

```
seq_documentation_block:
; Sequence 80, Application US/090826498
; Patent No. 6339068
; GENERAL INFORMATION:
;  APPLICANT:  Davis, Heather L.
;  APPLICANT:  Krieg, Arthur M.
;  APPLICANT:  Schorr, Joachim
;  APPLICANT:  Wu, Tong
;  TITLE OF INVENTION:  Vectors and Methods for Immunization or
;  TITLE OF INVENTION:  Therapeutic Protocols
;  FILE REFERENCE:  C1039/7009
;  CURRENT APPLICATION NUMBER: US/09/082,649B
;  CURRENT FILING DATE:  1998-05-20
;  PRIOR APPLICATION NUMBER:  US 60/047,233
;  PRIOR FILING DATE:  1997-05-20
;  PRIOR APPLICATION NUMBER:  US 60/047,209
;  PRIOR FILING DATE:  1997-05-20
;  NUMBER OF SEQ ID NOS:  85
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 80
;   LENGTH: 1360
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: plasmid DNA wild type Kanamycin resistance gene
US-09-082-649B-80


alignment_scores:
            Quality:    96.50               Length:     136
              Ratio:   1.322                  Gaps:       6
 Percent Similarity:  53.676   Percent Identity: 28.676

alignment_block:
 US-09-932-678-2 x US-09-082-649B-80   ..

 Align seq 1/1  to: US-09-082-649B-80  from: 1  to: 1360

  58 GluValLeuLeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLe 74
     |||::::::||||||:::     ||||||::::::     :::
 309 GAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAAT........... 347

  74 uLysAsnGlnLeuLeuAspProAspIleLysAspAspGlnIle...IleA 90
                             |||:::    |||:::::::::    :::|
 348 .........................GATGTTACAGATGACATGGTCAGACTAA 375

  90 sTrpLeuLeuGluPheArg......SerSerIleMetTyrLeuThrLys 104
     ||||||||    ||||||                 :::|||    ::::::    :::
 376 ATTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGT 425

 105 AspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnAr 121
              ::::::              :::              :::|||         :::
 426 ACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGAAAACAGC 475

 121 gSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValS 138
```

```
          |||   |||:::|||||||||              |||:::|||:
    476 ATTCCAG...GTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTG 522

    138 erAlaGlnThrValPheLeuArgProCysLeuSerMetIleAlaSerHis 154
        ::|||   :::|||||||||||         |||:::   :::   :::
    523 ATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAAT 572

    155 ......PheValProProArgValIle..................... 161
              |||         |||||||:::
    573 TGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACG 622

    162 .IleLysGluGlyAspValAspValSerAspSerAspAspGluAspAspA 178
         ::::::::::||   |||||||::::::::|   |||||||||   ::::
    623 AATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATG 672

    178 snLeuPro 180
        ::   |||
    673 GCTGGCCT 680

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-082-649B-81

seq_documentation_block:
; Sequence 81, Application US/09082649B
; Patent No. 6339068
; GENERAL INFORMATION:
;  APPLICANT:  Davis, Heather L.
;  APPLICANT:  Krieg, Arthur M.
;  APPLICANT:  Schorr, Joachim
;  APPLICANT:  Wu, Tong
;  TITLE OF INVENTION:  Vectors and Methods for Immunization or
;  TITLE OF INVENTION:  Therapeutic Protocols
;  FILE REFERENCE:  C1039/7009
;  CURRENT APPLICATION NUMBER: US/09/082,649B
;  CURRENT FILING DATE:  1998-05-20
;  PRIOR APPLICATION NUMBER:  US 60/047,233
;  PRIOR FILING DATE:  1997-05-20
;  PRIOR APPLICATION NUMBER:  US 60/047,209
;  PRIOR FILING DATE:  1997-05-20
;  NUMBER OF SEQ ID NOS:  85
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 81
;   LENGTH: 1360
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: plasmid DNA mutant Kanamycin resistance gene
US-09-082-649B-81


alignment_scores:
            Quality:   96.50            Length:    136
              Ratio:  1.322              Gaps:      6
 Percent Similarity:  53.676   Percent Identity: 28.676

alignment_block:
 US-09-932-678-2 x US-09-082-649B-81   ..

 Align seq 1/1  to: US-09-082-649B-81  from: 1  to: 1360

      58 GluValLeuLeuLysTyrLysLysGlyGluThrAsnAspPheGluLeuLe 74
         |||:::::|||||||:::   |||||:::::   :::
     309 GAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAAT.......... 347

      74 uLysAsnGlnLeuLeuAspProAspIleLysAspAspGlnIle...IleA 90
                              |||:::   |||:::::::::   :::|
     348 ......................GATGTTACAGATGAGATGGTCAGACTAA 375

      90 snTrpLeuLeuGluPheArg......SerSerIleMetTyrLeuThrLys 104
         |||||||   ||||||          :::|||   :::::   :::
     376 ACTGGCTGACAGAATTTATGCCTCTTCCCACCATCAAGCATTTTATACGT 425

     105 AspPheGluGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnAr 121
             ::::::         :::         ::::||      :::

     426 ACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCTGGAAAAACAGC 475

     121 qSerGlnThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValS 138
         |||   |||:::|||||||||                  |||:::|||:
     476 ATTCCAG...GTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTG 522

     138 erAlaGlnThrValPheLeuArgProCysLeuSerMetIleAlaSerHis 154
         ::|||   :::|||||||||||         |||:::   :::   :::
     523 ATGCGCTGGCAGTGTTCCTGAGACGTTTGCATTCGATTCCTGTTTGTAAT 572

     155 ......PheValProProArgValIle..................... 161
               |||         ||||||||:::
     573 TGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACG 622

     162 .IleLysGluGlyAspValAspValSerAspSerAspAspGluAspAspA 178
          ::::::::::|||   |||||||::::::::|   |||||||||   ::::
     623 AATGAATAATGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATG 672

     178 snLeuPro 180
         :   |||
     673 GCTGGCCT 680
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 02:33:51 ; Search time 2612.22 Seconds
(without alignments)
16566.781 Million cell updates/sec

Title: US-09-932-678-1
Perfect score: 2068
Sequence: 1 acagaggctgtggctggaag.....tgacatttgggattccccat 2068

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2068	100.0	2068	9	AF227156	AF227156 Homo sapi
2	2020	97.7	3756	9	HSA272050	AJ272050 Homo sapi
3	2018.4	97.6	2040	6	AX036047	AX036047 Sequence
4	1211.6	58.6	135873	2	AC012547	AC012547 Homo sapi
5	1211.6	58.6	192815	9	AC013444	AC013444 Homo sapi
6	567	27.4	2493	9	AK055742	AK055742 Homo sapi
7	506.4	24.5	1177	9	BC009198	BC009198 Homo sapi
8	420.8	20.3	1461	6	AX274907	AX274907 Sequence
9	420.8	20.3	2627	9	BC006441	BC006441 Homo sapi
10	208	10.1	129473	2	AC040158	AC040158 Homo sapi
c 11	208	10.1	143900	2	AC092137	AC092137 Homo sapi
12	192	9.3	175691	2	AC007615	AC007615 Homo sapi
13	192	9.3	191496	2	AC009130	AC009130 Homo sapi
14	192	9.3	208008	2	AC009093	AC009093 Homo sapi
15	190.8	9.2	187643	2	AC092562	AC092562 Papio ham
c 16	188.8	9.1	173166	9	AC092375	AC092375 Homo sapi
c 17	188.8	9.1	179150	2	AC017077	AC017077 Homo sapi
18	188.8	9.1	202004	9	HUAF001549	AF001549 Human Chr
19	187.2	9.1	170611	9	AC106788	AC106788 Homo sapi
20	158	7.6	143900	2	AC092137	AC092137 Homo sapi
c 21	153.2	7.4	191496	2	AC009130	AC009130 Homo sapi
c 22	153.2	7.4	196413	2	AC097269	AC097269 Pan troql
c 23	153.2	7.4	247381	2	AC097268	AC097268 Pan troql
24	146.4	7.1	2112	3	AY061123	AY061123 Drosophil
25	140.6	6.8	179150	2	AC017077	AC017077 Homo sapi
c 26	131	6.3	293	9	HS151B3F	256854 H.sapiens C
c 27	127	6.1	78347	2	AC022642	AC022642 Homo sapi
28	125.4	6.1	53814	2	AC107769	AC107769 Mus museu
c 29	118.6	5.7	53814	2	AC107769	AC107769 Mus museu
c 30	115	5.6	824	9	HSA343405	AJ343405 Homo sapi
31	108.6	5.3	821	9	HSA343406	AJ343406 Homo sapi
32	91.4	4.4	108040	2	AC068150	AC068150 Homo sapi
33	91.4	4.4	171930	2	AC092119	AC092119 Homo sapi
34	91.4	4.4	173166	9	AC092375	AC092375 Homo sapi
35	91.4	4.4	316296	2	AC092285	AC092285 Homo sapi
36	70.8	3.4	78347	2	AC022642	AC022642 Homo sapi
37	62.8	3.0	405	11	G27379	G27379 human STS S
38	62.8	3.0	6287	2	AC017809	AC017809 Drosophil
39	62.8	3.0	96018	3	AE003783	AE003783 Drosophil
40	62.8	3.0	172748	3	AC018484	AC018484 Drosophil
41	62.8	3.0	176424	3	AC011757	AC011757 Drosophil
42	59.6	2.9	174531	2	AC023752	AC023752 Drosophil
43	51.4	2.5	125020	9	AF429315	AF429315 Homo sapi
44	48.2	2.3	8797	6	AR027319	AR027319 Sequence
45	48.2	2.3	10338	4	BOVCASBE	M55158 Bovine beta

ALIGNMENTS

RESULT 1
AF227156
LOCUS AF227156 2068 bp mRNA linear PRI 29-APR-2000
DEFINITION Homo sapiens RRN3 mRNA, complete cds.
ACCESSION AF227156
VERSION AF227156.1 GI:7670099
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2068)
AUTHORS Moorefield,B., Greene,E.A. and Reeder,R.H.
TITLE RNA polymerase I transcription factor Rrn3 is functionally conserved between yeast and human
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (9), 4724-4729 (2000)
MEDLINE 20243763
REFERENCE 2 (bases 1 to 2068)
AUTHORS Moorefield,B., Greene,E.A. and Reeder,R.H.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2000) Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N., Seattle, WA 98109, USA
FEATURES Location/Qualifiers

```
     source          1..2068
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "16"
                     /map "16p12"
     CDS             69..2024
                     /note "similar to yeast RNA polymerase I transcription
                     factor Rrn3"
                     /codon_start 1
                     /product "RRN3"
                     /protein_id "AAF66160.1"
                     /db_xref "GI:7670100"
                     /translation "MAAPLLHTRLPGDAAASSSAVKKLGASRTGISNMRALENDFFNS
                     PPRKTVRFGGTVTEVLLKYKKGETNDFELLKNQLLDPDIKDDQIINWLLEFRSSIMYL
                     TKDFEQLISIILRLPWLNRSQTVVEEYLAFLGNLVSAQTVFLRPCLSMIASHFVPPRV
                     IIKEGDVDVSDSDDEDDNLPANFDTCHRALQIIARYVPSTPWFLMPILVEKFPFVRKS
                     ERTLECYVHNLLRISVYFPTLRHEILELIIEKLLKLDVNASRQGIEDAEETATQTCGG
                     TDSTEGLFNMDEDEETEHETKAGPERLDQMVHPVAERLDILMSLVLSYMKDVCYVDGK
                     VDNGKTKDLYRDLINIFDKLLLPTHASCHVQFFMFYLCSFKLGFAEAFLEHLWKKLQD
                     PSNPAIIRQAAGNYIGSFLARAKFIPLITVKSCLDLLVNWLHIYLNNQDSGTKAFCDV
                     ALHGPFYSACQAVFYTFVFRHKQLLSGNLKEGLQYLQSLNFERIVMSQLNPLKICLPS
                     VVNFFAAITNKYQLVFCYTIIERNNRQMLPVIRSTAGGDSVQICTNPLDTFFPFDPCV
                     LKRSKKFIDPIYQVWEDMSAEELQEFKKPMKKDIVEDEDDDFLKGEVPQNDTVIGITP
                     SSFDTHFRSPSSSVGSPPVLYMQPSPL"
BASE COUNT      573 a    437 c    485 g    573 t
ORIGIN

  Query Match             100.0%;  Score 2068;  DB 9;  Length 2068;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2068;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 acagaggctgtggctggaaggagctgggcatccggcctgaggcgcagcggtcgcgttagt 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACAGAGGCTGTGGCTGGAAGGAGCTGGGCATCCGGCCTGAGGCGCAGCGGTCGCGTTAGT 60

Qy         61 tcggcccaatggcggcaccgctgcttcacacgcgtttgccgggagatgcggccgcttcgt 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCGGCCCAATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGATGCGGCCGCTTCGT 120

Qy        121 cctctgcagttaagaagctgggcgcgtcgaggactgggatttcaaatatgcgtgcattag 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATATGCGTGCATTAG 180

Qy        181 agaatgacttttttcaattctcccccaagaaaaactgttcggtttggtggaactgtgacag 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTTCGGTTTGGTGGAACTGTGACAG 240

Qy        241 aagtcttgctgaagtacaaaaagggtgaaacaaatgactttgagttgttgaagaaccagc 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGC 300

Qy        301 tgttagatccagacataaaggatgaccagatcatcaactggctgctagaattccgttctt 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGTTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTT 360

Qy        361 ctatcatgtacttgacaaaagactttgagcaacttatcagtattatattaagattgcctt 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCCTT 420

Qy        421 ggttgaatagaagtcaaacagtagtggaagagtatttggcttttcttggtaatcttgtat 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTGGTAATCTTGTAT 480

Qy        481 cagcacagactgttttcctcagaccgtgtctcagcatgattgcttcccattttgtgcctc 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTTGTGCCTC 540

Qy        541 cccgagtgatcattaaggaaggcgatgtagatgtttcagattctgatgatgaagatgata 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGATGATGAAGATGATA 600

Qy        601 atcttcctgcaaattttgacacatgtcacagagccttgcaaataatagcaagatatgtac 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATCTTCCTGCAAATTTTGACACATGTCACAGAGCCTTGCAAATAATAGCAAGATATGTAC 660

Qy        661 catcgacaccgtggtttctcatgccaatactggtggaaaaatttccatttgttcgaaaat 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTCGAAAAT 720

Qy        721 cagagagaacactggaatgttacgttcataacttactaaggattagtgtatattttccaa 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CAGAGAGAACACTGGAATGTTACGTTCATAACTTACTAAGGATTAGTGTATATTTTCCAA 780

Qy        781 ccttgaggcatgaaattctggagcttattattgaaaaactactcaagttggatgtgaatg 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAAACTACTCAAGTTGGATGTGAATG 840

Qy        841 catcccggcagggtattgaagatgctgaagaaacagcaactcaaacttgtggtgggacag 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CATCCCGGCAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAACTTGTGGTGGGACAG 900

Qy        901 attccacggaaggattgtttaatatggatgaagatgaagaaactgaacatgaaacaaagg 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATTCCACGGAAGGATTGTTTAATATGGATGAAGATGAAGAAACTGAACATGAAACAAAGG 960

Qy        961 ctggtcctgaacggctcgaccagatggtgcatcctgtagccgagcgcctggacatcctga 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGGTCCTGAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCGCCTGGACATCCTGA 1020

Qy       1021 tgtctttggtttttgtcctacatgaaggatgtctgctatgtagatggtaaggttgataacg 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TGTCTTTGGTTTTTGTCCTACATGAAGGATGTCTGCTATGTAGATGGTAAGGTTGATAACG 1080

Qy       1081 gcaaaacaaaggatctatatcgcgacctgataaacatctttgacaaactcctgttgccca 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GCAAAACAAAGGATCTATATCGCGACCTGATAAACATCTTTGACAAACTCCTGTTGCCCA 1140

Qy       1141 cccatgcctcctgccatgtacagtttttcatgttttacctctgtagtttcaaattgggat 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCATGCCTCCTGCCATGTACAGTTTTTCATGTTTTACCTCTGTAGTTTCAAATTGGGAT 1200

Qy       1201 tcgcagaggcatttttggaacatctctggaaaaaattgcaggacccaagtaatcctgcca 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGCAGAGGCATTTTTGGAACATCTCTGGAAAAAATTGCAGGACCCAAGTAATCCTGCCA 1260

Qy       1261 tcatcaggcaggctgctggaaattatattggaagctttttggcaagagctaaatttattc 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAGAGCTAAATTTATTC 1320

Qy       1321 ctcttattactgtaaaatcatgcctagatcttttggttaactggctgcacatataccttа 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTCTTATTACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTA 1380

Qy       1381 ataaccaggattcgggaacaaaggcattctgcgatgttgctctccatggaccatttttact 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 ATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACT 1440

Qy       1441 cagcctgccaagctgtgttctacacctttgtttttagacacaagcagctttgagcggaa 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTGAGCGGAA 1500

Qy       1501 acctgaaagaaggtttgcagtatcttcagagtctgaatttttgagcggatagtgatgagcc 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 ACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCC 1560

Qy       1561 agctaaatcccctgaagatttgcctgccctcagtggttaactttttttgctgcaatcacaa 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 AGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGCAATCACAA 1620

Qy       1621 ataagtaccagctcgtcttctgctacaccatcattgagaggaacaatcgccagatgctgc 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 ATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAATCGCCAGATGCTGC 1680

Qy       1681 cagtcattaggagtaccgctggaggagactcagtgcagatctgcacaaacccgctggaca 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    1681 CAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCACAAACCCGCTGGACA 1740

Qy    1741 ccttcttccccttttgatccctgtgtgctgaagaggtcaaagaaattcattgatcctattt 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 CCTTCTTCCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATTCATTGATCCTATTT 1800

Qy    1801 atcaggtatgggaagacatgagtgctgaagagctacaggagttcaagaaacccatgaaaa 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 ATCAGGTATGGGAAGACATGAGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAA 1860

Qy    1861 aggacatagtggaagatgaagatgatgacttttctgaaaggcgaagtgccccagaatgata 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 AGGACATAGTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCAGAATGATA 1920

Qy    1921 ccgtgattgggatcacaccaaagctccttgacacgcatttccgaagtccttcaagtagtg 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 CCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAGTCCTTCAAGTAGTG 1980

Qy    1981 tgggctcccccacccgtgttgtacatgcaacccagtcccctctgacggcagaaatttgtga 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 TGGGCTCCCCACCCGTGTTGTACATGCAACCCAGTCCCCTCTGACGGCAGAAATTTGTGA 2040

Qy    2041 ctgagatgtgacatttgggattccccat 2068
           |||||||||||||||||||||||||||||
Db    2041 CTGAGATGTGACATTTGGGATTCCCCAT 2068


RESULT   2
HSA272050
LOCUS       HSA272050               3756 bp    mRNA    linear   PRI 07-SEP-2000
DEFINITION  Homo sapiens mRNA for transcription initiation factor IA protein
            (TIF-IA gene).
ACCESSION   AJ272050
VERSION     AJ272050.1  GI:10046713
KEYWORDS    TIF-IA; transcription initiation factor.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3756)
  AUTHORS   Bodem,J., Hoffmann-Rohrer,U., Ross,W., Delius,H., Vingron,M. and
            Grummt,I.
  TITLE     Cloning and functional characterization of transcription initiation
            factor TIF-IA, a growth-dependent regulator of ribosomal RNA
            synthesis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3756)
  AUTHORS   Bodem,J.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2000) Bodem J.J., Molekulare Biologie der Zelle
            II, Deutsches Krebsforschungszentrum, INF.280, 69120 Heidelberg,
            GERMANY
FEATURES             Location/Qualifiers
     source          1. .3756
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            23. .3756
                     /gene="TIF-IA"
     CDS             23. .1978
                     /gene="TIF-IA"
                     /function="transcription initiation factor"
                     /codon_start=1
                     /evidence=experimental
                     /product="transcription initiation factor IA protein"
                     /protein_id="CAC07955.1"
                     /db_xref="GI:10046714"
                     /translation="MAAPLLHTRLPGDAAASSSAVKKLGASRTGISNMRALENDFFNS
                     PPRKTVRFGGTVTEVLLKYKKGETNDFELLKNQLLDPDIKDDQIINWLLEFRSSIMYL
                     TKDFEQLISIILRLPWLNRSQTVVEEYLAFLGNLVSAQTVFLRPCLSMIASHFVPPRV
                     IIKEGDVDVSDSDDEDDNLPANFDTCHRALQIIARYVPSTPWFLMPILVEKFPFVRKS
                     ERTLECYVHNLLRISVYFPTLRHEILELIIEKLLKLDVNASRQGIEDAEETATQTCGG
                     TDSTEGLFNMDEDEETEHETKAGPERLDQMVHPVAERLDILMSLVLSYMKDVCYVDGK
                     VDNGKTKDLYRDLINIFDKLLLPTHASCHVQFFMFYLCSFKLGFAEAFLEHLWKKLQD
```

```
                     PSNPAIIRQAAGNYIGSFLARAKFIPLITVKSCLDLLVNWLHIYLNNQDSGTKAPCDV
                     ALHGPFYSACQAVFYXFVFRHKQLLSGNLKEGLQYLQSLNFEEIVMSQLNFLKICLFS
                     VVNFFAAITNKYQLVFCYTIIERNNRQMLPVIRSTAGGDSVQICINPLDTFFPFDPCV
                     LKRSKKFIDPIYQVWEDMSAEELQEFKKPMKKDIVEDEDDDFLKGEVPQNDIVIGTIP
                     SSFDTHFRSPSSSVGSPPVLYMQPSPL"
     3'UTR           1979. .3756
                     /gene="TIF-IA"
                     /evidence=experimental
BASE COUNT     1060 a     692 c     844 g    1156 t       4 others
ORIGIN


  Query Match           97.7%;  Score 2020;  DB 9;  Length 3756;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2020;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      47 gcggtcgcgttagttcggcccaatggcggcaccgctgcttcacacgcgtttgccgggaga 106
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCGGTCGCGTTAGTTCGGCCCAATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGA 60

Qy     107 tgcggccgcttcgtcctctgcagttaagaagctgggcggcgtcgaggactgggatttcaaa 166
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGCGGCCGCTTCGTCCTCTGCAGTTAAGAAGCTGGGCGGCGTCGAGGACTGGGATTTCAAA 120

Qy     167 tatgcgtgcattaagaagaatgacttttcaattctccccccaaggaaaaactgttcggtttgg 226
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TATGCGTGCATTAGAGAATGACTTTTCAATTCTCCCCCAAGAAAAACTGTTCGGTTTGG 180

Qy     227 tggaactgtgacagaagtcttgctgaagtacaaaaagggtgaaacaaatgacttggagtt 286
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTGAGTT 240

Qy     287 attgaagaaccagctgttagatccagacataaagggatgaccagatcatcaactggctggct 346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTTGAAGAACCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGGCT 300

Qy     347 agaattccgttcttctatcatgtacttgacaaaagactttgagcaacttatcagtattat 406
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGAATTCCGTTCTTCTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTAT 360

Qy     407 attaagattgccttggttgaatagaagtcaaacagtagtggaagagtatttggcttttct 466
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATTAAGATTGCCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCT 420

Qy     467 tggtaatcttgtatcagcacagactgttttcctcagaccgtgtctcagcatgattgcttc 526
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTC 480

Qy     527 ccattttgtgcctccccgagtgatcattaaggaaggcgatgtagatgtttcagattctga 586
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGA 540

Qy     587 tgatgaagatgataatcttcctgcaaatttttgacacatgtcacagaagccttgcaaataat 646
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TGATGAAGATGATAATCTTCCTGCAAATTTTGACACATGTCACAGAGCCTTGCAAATAAT 600

Qy     647 agcaagatatgtaccatcgacaccgtggtttctcatgccaatactggtggaaaaatttcc 706
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCC 660

Qy     707 atttgttcgaaaatcagagagaacactggaatgttacgttcataacttactaaggattag 766
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 ATTTGTTCGAAAATCAGAGAGAACACTGGAATGTTACGTTCATAACTTACTAAGGATTAG 720

Qy     767 tgtatattttccaacctttgaggcatgaaattctggagcttattattgaaaaactactcaa 826
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 TGTATATTTTCCAACCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAAACTACTCAA 780

Qy     827 gttggatgtgaaatgcatcccggcagggtattgaagatgctgaagaaacagcaactctaaac 886
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GTTGGATGTGAAATGCATCCCGGCAGGGTATTGAAGATGCTGAAGAAACAGCAACTCTAAAC 840

Qy     887 ttgtggtggacagattccacggaaggattgtttaatatggatgaagatgaagaaactga 946
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTGTGGTGGGACAGATTCCACGGAAGGATTGTTTAATATGGATGAAGATGAAGAAACTGA 900

Qy        947 acatgaaacaaaggctggtcctgaacggctcgaccagatggtgcatcctgtagccgagcg 1006
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACATGAAACAAAGGCTGGTCCTGAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCG 960

Qy       1007 cctggacatcctgatgtctttggttttgtcctacatgaaggatgtctgctatgtagatgg 1066
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCTGGACATCCTGATGTCTTTGGTTTTGTCCTACATGAAGGATGTCTGCTATGTAGATGG 1020

Qy       1067 taaggttgataacggcaaaacaaaggatctatatcgcgacctgataaacatctttgacaa 1126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TAAGGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATCTTTGACAA 1080

Qy       1127 actcctgttgcccacccatgcctcctgccatgtacagttttcatgttttacctctgtag 1186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACTCCTGTTGCCCACCCATGCCTCCTGCCATGTACAGTTTTCATGTTTTACCTCTGTAG 1140

Qy       1187 tttcaaattgggattcgcagaggcatttttggaacatctctggaaaaaattgcaggaccc 1246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TTTCAAATTGGGATTCGCAGAGGCATTTTTGGAACATCTCTGGAAAAAATTGCAGGACCC 1200

Qy       1247 aagtaatcctgccatcatcaggcaggctgctggaaattatattggaagctttttggcaag 1306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAG 1260

Qy       1307 agctaaatttattcctcttattactgtaaaatcatgcctagatcttttggttaactggct 1366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AGCTAAATTTATTCCTCTTATTACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCT 1320

Qy       1367 gcacatatacctttaataaccaggattcgggaacaaaggcattctgcgatgttgctctcca 1426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GCACATATACCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCA 1380

Qy       1427 tggaccattttactcagcctgccaagctgtgttctacacctttgtttttagacacaagca 1486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACRCCTTTGTTTTTAGACACAAGCA 1440

Qy       1487 gcttttgagcggaaacctgaaagaaggtttgcagtatcttcagagtctgaattttgagcg 1546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GCTTTTGAGCGGAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCG 1500

Qy       1547 gatagtgatgagccagctaaatcccctgaagatttgcctgccctcagtggttaacttttt 1606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTT 1560

Qy       1607 tgctgcaatcacaaataagtaccagctcgtcttctgctacaccatcattgagaggaacaa 1666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TGCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAA 1620

Qy       1667 tcgccagatgctgccagtcattaggagtaccgctggaggagactcagtgcagatctgcac 1726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TCGCCAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCAC 1680

Qy       1727 aaacccgctggacaccttcttcccctttgatccctgtgtgctgaagaggtcaaagaaatt 1786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 AAACCCGCTGGACACCTTCTTCCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATT 1740

Qy       1787 cattgatcctatttatcaggtatgggaagacatgagtgctgaagagctacaggagttcaa 1846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CATTGATCCTATTTATCAGGTCTGGGAAGACATGAGTGCTGAAGAGCTACAGGAGTTCAA 1800

Qy       1847 gaaacccatgaaaaaggacatagtggaagatgaagatgatgactttctgaaaggcgaagt 1906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GAAACCCATGAAAAAGGACATAGTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGT 1860

Qy       1907 gccccagaatgataccgtgattgggatcacaccaagctcctttgacacgcatttccgaag 1966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GCCCCAGAATGATACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAG 1920

Qy       1967 tccttcaagtagtgtgggctccccacccgtgttgtacatgcaacccagtcccctctgacg 2026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TCCTTCAAGTAGTGTGGGCTCCCCACCCGTGTTGTACATGCAACCCAGTCCCCTCTGACG 1980

Qy       2027 gcagaaatttgtgactgagatgtgacatttgggattccccat 2068
              ||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCAGAAATTTGTGACTGAGATGTGACATTTGGGATTCCCCAT 2022


RESULT      3
AX036047
LOCUS       AX036047                2040 bp    DNA     linear   PAT 15 NOV 2000
DEFINITION  Sequence 1 from Patent WO0055316.
ACCESSION   AX036047
VERSION     AX036047.1  GI:11191586
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2040)
  AUTHORS   Grummt,I. and Vingron,M.
  TITLE     Rna polymerase i transcription factor tif ia
  JOURNAL   Patent: WO 0055316-A 1 21 SEP 2000;
            DEUTSCHES KREBSFORSCH (DE) ; GRUMMT INGRID (DE) ; VINGRON MARTIN
            (DE)
FEATURES             Location/Qualifiers
     source          1..2040
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             40..1995
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC16268.1"
                     /db_xref="GI:11191587"
                     /translation="MAAPLLHTRLPGDAAASSSAVKKLGASRTGISNMRALENDFFNS
                     PPRKTVRFGGTVTEVLLKYKKGETNDFELLKNQLLDPDIKDDQIINWLLEFRSSIMYL
                     TKDFEQLISIILRLPWLNRSQIVVEEYLAFLGNLVSAQTVFLRPCLSMIASHFVPPRV
                     IIKEGDVDVSDSDDEDDNLPANFDTCHRALQIIARYVPSTPWFLMPILVEKFPFVRKS
                     ERTLECYVHXLLRISVYFPTLRHEILELIIEKLLKLDVNASRQGIEDAEETATQTCGG
                     TDSTEGLFNMDEDEETEHETKAGPERLDQMVHPVAERLDILMSLVLSYMKDVCYVDGK
                     VDNGKTKDLYRDLINIFDKLLLPTHASCHVQFFMFYLCSFKLGFAEAFLEHLWKKLQD
                     PSNPAIIRQAAGNYIGSFLARAKFIPLITVKSCLDLLVNWLHIYLNNQDSGTKAFCDV
                     ALHGPFYSACQAVFYXFVFRHKQLLSGNLKEGLQYLQSLNFERIVMSQLNPLKICLPS
                     VVNFFAAITNKYQLVFCYTIIERNNRQMLPVIRSTAGGDSVQICTNPLDTFFPFDPCV
                     LKRSKKFIDPIYQVWEDMSAEELQEFKKPMKKDIVEDEDDDFLKGEVPQNDTVIGITP
                     SSFDTHFRSPSSSVGSPPVLYMQPSPL"
BASE COUNT      563 a    433 c    474 g    568 t        2 others
ORIGIN


  Query Match              97.6%;  Score 2018.4;  DB 6;  Length 2040;
  Best Local Similarity    99.9%;  Pred. No. 0;
  Matches 2019;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         47 gcggtcgcgttagttcggcccaatggcggcaccgctgcttcacacgcgtttgccgggaga 106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         18 GCGGTCGCGTTAGTTCGGCCCAATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGA 77

Qy        107 tgcggccgcttcgtcctctgcagttaagaagctgggcgcgtcgaggactgggatttcaaa 166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 TGCGGCCGCTTCGTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAA 137

Qy        167 tatgcgtgcattagagaatgactttttcaattctcccccaagaaaaactgttcggtttgg 226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 TATGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTTCGGTTTGG 197

Qy        227 tggaactgtgacagaagtcttgctgaagtacaaaaagggtgaaacaaatgactttgagtt 286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 TGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTT 257

Qy        287 gttgaagaaccagctgttagatccagacataaaggatgaccagatcatcaactggctgct 346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 GTTGAAGAACCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGCT 317
```

```
Qy     347 agaattccgttcttctatcatgtacttgacaaaagactttgagcaacttatcagtattat 406
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     318 AGAATTCCGTTCTTCTATCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTAT 377

Qy     407 attaagattgccttggttgaatagaagtcaaacagtagtggaagagtatttggctttttct 466
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     378 ATTAAGATTGCCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCT 437

Qy     467 tggtaatcttgtatcagcacagactgttttcctcagaccgtgtctcagcatgattgcttc 526
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     438 TGGTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTGCTTC 497

Qy     527 ccatttttgtgcctccccgagtgatcattaaggaaggcgatgtagatgtttcagattctga 586
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     498 CCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGA 557

Qy     587 tgatgaagatgataatcttcctgcaaattttgacacatgtcacagaagccttgcaaataat 646
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     558 TGATGAAGATGATAATCTTCCTGCAAATTTTGACACATGTCACAGAGCCTTGCAAATAAT 617

Qy     647 agcaagatatgtaccatcgacaccgtggtttctcatgccaatactggtggaaaaatttcc 706
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     618 AGCAAGATATGTACCATCGACACCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCC 677

Qy     707 atttgttcgaaaatcagagagaacactggaatgttacgttcataacttactaaggattag 766
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     678 ATTTGTTCGAAAATCAGAGAGAACACTGGAATGTTACGTTCATNACTTACTAAGGATTAG 737

Qy     767 tgtatatttttccaaccttgaggcatgaaattctggagcttattattgaaaaactactcaa 826
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     738 TGTATATTTTCCAACCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAAACTACTCAA 797

Qy     827 gttggatgtgaatgcatcccggcagggtattgaagatgctgaagaaacagcaactcaaac 886
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     798 GTTGGATGTGAATGCATCCCGGCAGGGTATTGAAGATGCTGAAGAAACAGCAACTCAAAC 857

Qy     887 ttgtggtgggacagattccacggaaggattgtttaatatggatgaagatgaagaaactga 946
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     858 TTGTGGTGGGACAGATTCCACGGAAGGATTGTTTAATATGGATGAAGATGAAGAAACTGA 917

Qy     947 acatgaaacaaaggctggtcctgaacggctcgaccagatggtgcatcctgtagccgagcg 1006
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     918 ACATGAAACAAAGGCTGGTCCTGAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCG 977

Qy    1007 cctggacatcctgatgtctttggttttgtcctacatgaaggatgtctgctatgtagatgg 1066
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     978 CCTGGACATCCTGATGTCTTTGGTTTTGTCCTACATGAAGGATGTCTGCTATGTAGATGG 1037

Qy    1067 taaggttgataacggcaaaacaaaggatctatatcgcgacctgataaacatctttgacaa 1126
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1038 TAAGGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATCTTTGACAA 1097

Qy    1127 actcctgttgcccacccatgcctcctgccatgtacagtttttcatgttttacctctgtag 1186
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1098 ACTCCTGTTGCCCACCCATGCCTCCTGCCATGTACAGTTTTTCATGTTTTACCTCTGTAG 1157

Qy    1187 tttcaaattgggattcgcagaggcatttttggaacatctctggaaaaaaattgcaggaccc 1246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1158 TTTCAAATTGGGATTCGCAGAGGCATTTTTGGAACATCTCTGGAAAAAATTGCAGGACCC 1217

Qy    1247 aagtaatcctgccatcatcaggcaggctgctggaaattatattggaagctttttggcaag 1306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1218 AAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAG 1277

Qy    1307 agctaaatttattcctcttattactgtaaaatcatgcctagatcttttggttaactggct 1366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1278 AGCTAAATTTATTCCTCTTATTACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCT 1337

Qy    1367 gcacatatacccttaataaccaggattcgggaacaaaggcattctgcgatgttgctctcca 1426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1338 GCACATATACCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCA 1397

Qy    1427 tggaccatttactcagcctgccaagctgtgttctacacctttgttttagacacaagca 1486
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1398 TGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACNCCTTGTTTTTAGACACAAGCA 1457

Qy    1487 gcttttgaagcggaaaacctgaaagaaggtttgcagtatcttcagagtctgaatttgaagcg 1546
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1458 GCTTTTGAGCGGAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTGAGCG 1517

Qy    1547 gatagtgatgaagccagctaaatcccctgaagatttgcctgcccctcagtggttaacttttt 1606
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1518 GATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTT 1577

Qy    1607 tgctgcaatcacaaataagtaccagctcgtcttctgctacaccatcattgagaggaacaa 1666
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1578 TGCTGCAATCACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAA 1637

Qy    1667 tcgccagatgctgccagtcattaggagtaccgctggaggagactcagtgcagatctgcac 1726
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1638 TCGCCAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAGATCTGCAC 1697

Qy    1727 aaaccccgctggacaccttcttccccttttgatccctgtgtgctgaaggagtcaaaggaatt 1786
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1698 AAACCCGCTGGACACCTTCTTCCCCTTTGATCCCTGTGTGCTGAAGAGGTCAAAGAAATT 1757

Qy    1787 gattgatcctatttatcaggtatgggaagacatgagtgctgaagaggctacaggagttcaa 1846
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1758 GATTGATCCTATTTATCAGGTGTGGGAAGACATGAGTGCTGAAGAGCTACAGGAGTTCAA 1817

Qy    1847 gaaacccatgaaaaaggacatagtgaaagatgaagatgatgacttttctgaaaggccgaagt 1906
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1818 GAAACCCATGAAAAAGGACATAGTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGT 1877

Qy    1907 gccccagaatgatacccgtgattgggatcacaccaagctccttgacacgcatttccgaag 1966
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1878 GCCCCAGAATGATACCGTGATTGGGATCACACCAAGCTCCTTGACACGCATTTCCGAAG 1937

Qy    1967 tccttcaagtagtgtgggctcccacccgtgttgtacatgcaacccagtcccctctgacg 2026
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1938 TCCTTCAAGTAGTGTGGGCTCCCACCCGTGTTGTACATGCAACCCAGTCCCCTCTGACG 1997

Qy    2027 gcagaaatttgtgactgagatgtgacatttgggattccccat 2068
           |||||||||||||||||||||||||||||||||||||||||||
Db    1998 GCAGAAATTTGTGACTGAGATGTGACATTTGGGATTCCCCAT 2039
```

RESULT 4
AC012547
LOCUS AC012547 135873 bp DNA linear HTG 30-MAR-2000
DEFINITION Homo sapiens clone RP11-11K1, WORKING DRAFT SEQUENCE, 14 unordered pieces.
ACCESSION AC012547
VERSION AC012547.2 GI:7341888
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 135873)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-11K1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 135873)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Haqos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C. Kann,L., Karatas,A., Klein,J., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

```
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 30, 2000 this sequence version replaced gi:6139119.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
                        Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
                  -------- Project Information
               Center project name: L3178
               Center clone name: 11_K_1
                  ---      Summary Statistics
               Sequencing vector: M13; M77815; 100% of reads
               Chemistry: Dye terminator Big Dye; 100% of reads
               Assembly program: Phrap; version 0.960731
               Consensus quality: 127596 bases at least Q40
               Consensus quality: 130929 bases at least Q30
               Consensus quality: 132388 bases at least Q20
               Insert size: 133000; agarose-fp
               Insert size: 134573; sum-of-contigs
               Quality coverage: 4.9 in Q20 bases; agarose-fp
               Quality coverage: 4.9 in Q20 bases; sum-of-contigs
                  ---------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      1516: contig of 1516 bp in length
            *     1517  1616: gap of       100 bp
            *     1617      3136: contig of 1520 bp in length
            *     3137  3236: gap of       100 bp
            *     3237      5148: contig of 1912 bp in length
            *     5149  5248: gap of       100 bp
            *     5249      7742: contig of 2494 bp in length
            *     7743  7842: gap of       100 bp
            *     7843     10650: contig of 2808 bp in length
            *    10651 10750: gap of       100 bp
            *    10751     16194: contig of 5444 bp in length
            *    16195 16294: gap of       100 bp
            *    16295     25751: contig of 9457 bp in length
            *    25752 25851: gap of       100 bp
            *    25852     37209: contig of 11358 bp in length
            *    37210 37309: gap of       100 bp
            *    37310     47828: contig of 10519 bp in length
            *    47829 47928: gap of       100 bp
            *    47929     64239: contig of 16311 bp in length
            *    64240 64339: gap of       100 bp
            *    64340     77846: contig of 13507 bp in length
            *    77847 77946: gap of       100 bp
            *    77947     95311: contig of 17365 bp in length
            *    95312 95411: gap of       100 bp
            *    95412    108930: contig of 13519 bp in length
            *   108931 109030: gap of       100 bp
            *   109031    135873: contig of 26843 bp in length.
FEATURES             Location/Qualifiers
     source          1..135873
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-11K1"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1516
                     /note="assembly_fragment"
     misc_feature    1617..3136
                     /note="assembly_fragment"
     misc_feature    3237..5148
                     /note="assembly_fragment"
     misc_feature    5249..7742
                     /note="assembly_fragment"
     misc_feature    7843..10650
                     /note="assembly_fragment"
     misc_feature    10751..16194
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    16295..25751
                     /note="assembly_fragment"
     misc_feature    25852..37209
                     /note="assembly_fragment"
     misc_feature    37310..47828
                     /note="assembly_fragment"
     misc_feature    47929..64239
                     /note="assembly_fragment"
     misc_feature    64340..77846
                     /note="assembly_fragment"
     misc_feature    77947..95311
                     /note="assembly_fragment"
     misc_feature    95412..108930
                     /note="assembly_fragment"
     misc_feature    109031..135873
                     /note="assembly_fragment"
BASE COUNT    42279 a  25994 c  25146 g  41147 t   1307 others
ORIGIN

  Query Match             58.6%;  Score 1211.6;  DB 2;  Length 135873;
  Best Local Similarity   80.2%;  Pred. No. 0;
  Matches 1617;  Conservative    0;  Mismatches  334;  Indels   65;  Gaps   14;

Qy         60 ttcggcccaatggcggcaccgctgcttcacacgcgtttgccgggagatgcggccgcttcg 119
              ||  ||||    |||   |     |||||||| |||    | |||||||| || || |   |||||
Db      96731 TTAAGCCCCGGGGCAACCAGGCTGCTTCTCACATGGTTGCCGGG-GAAGCAGAGGCTTCA 96789

Qy        120 tcctctgcagttaagaagctgggcgcgtcgaggactgggatttcaaatatgcgtgcatta 179
                | ||||   || ||||||||||     |||   ||||    | |||||| |||||| ||||||
Db      96790 GCTTCTGTGGTCAAGAAGCTGGGTATGTCATGGACCAGTATTTCAGATATGCTTGCATTC 96849

Qy        180 gagaatgacttttcaattctcccccaagaaaaactgttcggtttggtggaactgtgaca 239
                ||||||| |||| ||   |||||||||||||||| |||||     ||| |||||| | ||||
Db      96850 AAGAATGATTTTTCCAGCTCTCCCCCAAGAAAAATTGTTCATCTTGATGGAACGGGGACA 96909

Qy        240 gaagtcttgctgaagtacaaaaagggtgaaacaaatgactttgagttgttgaaggaccag 299
              |||||||||| |||| ||||| || |    | |  |   | |||||||| ||||||||| ||||
Db      96910 GAAGTCTTGCCGAAGGACAAAGAGTGAAACAAATTTAACTTTGAGCTGTTGAAGAGCCAG 96969

Qy        300 ttgttagatccagacataaaggatgaccagatcatcaactggctgctagaattccgttct 359
                ||| |||| || |||||||||||||||||||| | |||||||||    | ||||   | |||
Db      96970 TTGTGAGATTCAAACATAAAGGATGACCAGATCTTGAACTGGCTGTGAAAATTTTGCTCT 97029

Qy        360 tctatcatgtacttgacaaaagactttgagcaacttatcagtattatattaagattgcct 419
              |||||||||||||||| ||||||||||||||||| ||   |||||||||| ||| ||||||||||
Db      97030 TCTATCATGTACTTGGCAAAAGACTTTGAGCAGCTCCTCAGTATTACATTGAGATTGCCT 97089

Qy        420 tggttgaatagaaagtcaaacagtagtggaaagagtatttggctttcttggtaatcttgta 479
              |||||||||||||| |||||||||| ||||||||||||||||||||||| ||| |||| |||||||
Db      97090 TGGTTGAATAGAATTCAAACAGTGGTGGAAGAGTATTTGGCTTCTCTGGGTAGTCTTGTA 97149

Qy        480 ccagcacagactgttttcctcagaccgtgtctcagcatgattgcttcccattttgtg    536
                || || |         | ||||| || ||  ||||||||||| |||||||||||||| | |
Db      97150 ACAACAGA----CTCTTCCTTAGCCCAAGTCTCAGCATAATTGCTTCCCATTTGGCGTAC 97205

Qy        537 cctccccgagtgatcattaaggaaggcgatgtagatgtttcagattctgatgatgaagat 596
              |||||   || |||||||||||| |    || ||||| ||||||||||||||||||||| ||||
Db      97206 CCTCCTTGAATGATCATTAAGAAGAGCCATGTAAATGTTTCAGATTCTGATGATGGAGAT 97265

Qy        597 gataatcttcctgcaaattttgacacatgtcacagagccttgcaaataatagcaagatat 656
              ||| |||||||||| |||||||||| |||| | | |||| ||||||| |||||||||
```

```
Db  97266 GATTATCTTCCTGCAGATTTTGACACACGTCATACAACCTTACAAATAACAGCAAGATAT 97325

Qy    657 gtaccatcgacaccgtggtttctcatgccaatactggtggaaaaatttccatttgttcga 716
          || ||| | ||| ||||||||| |||| ||||||||||||| ||||||||||||||
Db  97326 GTCCCACTGGCACTGTGGTTTCTTTTGCCCATACTGGTGGAAACATTTCCATTTGTTC-- 97383

Qy    717 aaatcagaagaagaacactggaatgttacgttcataacttactaaggattagtgtatatttt 776
          |||||||||||||||||||||||||| |||||||||||||| ||||||| |||| |||
Db  97384 -AATCAGAGAGAACACTGGAATGTTACATTCATAACTTACTAAAGATTAGTATATACTTT 97442

Qy    777 ccaaccttgaggcatgaaatttctggagcttattattgaaaaactactcaagttggatgtg 836
          |||||||||||| |||||||||| |||||||||| |||| ||| |||| |||||||||
Db  97443 CCAACCTTGAGGAATGAAATTCTGAAGCTTATTATTAAAAATCTATTCAAATTGGATGTG 97502

Qy    837 aatgcatcccggcaggtattgaagatgctgaagaaacagcaactcaaacttgtggtggg 896
          ||| ||| | |||| | |||||||||||||| |||||||| ||||||||||||||||
Db  97503 AATACATTCTGGCATGATATTGAAGATGCTGAGGAAACAGCAGCTCAAACTTGTGGTGGG 97562

Qy    897 acagattccacggaaggattgtttaatatggatgaagatgaagaaactgaacatgaaaca 956
          |||||||||| |||| |||||| |||||| | |||||||||| ||      ||   |
Db  97563 ACAGATTCCATGGAATGATTGTTCAATATGAACGAAGATGAAGAGAC------TGTTTCT 97616

Qy    957 aaggctggtcctgaacggctcgaccagatggtgcatcctgtagccgaagcgcctgggacatc 1016
          |||||||   |  || ||||||  |||||||||| |||||||| || ||||||||| |
Db  97617 AAGGCTGACTCCAAATGGCTCGGTCAGATGGTGCACCCTGTAGCTGAATGCCTGGACACC 97676

Qy   1017 ctgatgtctttggttttgtcctacatgaaggatgtctgctatg-----tagatggtaaggg 1071
          ||| ||||||||| |||||   ||| ||||| |||||||||||     || |||||||||
Db  97677 CTGCTGTCTTTGCTTTTGCAGTACTTGAAGTATGTCTGCTATGTAAAATACATGGTAAGG 97736

Qy   1072 ttgataacggcaaaacaaaggatctatatcgcgaccctgataaacatctttgacaaactcc 1131
          |||||||| ||||| ||||||| ||||      |||          ||| ||||
Db  97737 TTGATAACAACAAAATAAAGGATTTATA--------ATATATCACAAACTGATAAACATA 97788

Qy   1132 tgttgcccacccatgcctcctgccatgtacagtttttcatgttttacctctgtagtttca 1191
          | |  |  |||| |||||||| || ||||||||| |||||||||||||| | || |
Db  97789 TTTGACAAACCCCTGCCTCCTGTCACGTACAGTTTTTTATGTTTTACCTCTGTGGCTTAA 97848

Qy   1192 aattgggattcgcagaggcatttttgggaacatctctgg---aaaaaattgcaggacccaa 1248
          ||||| || || |||| |||||||||||||| |||||||    |||||| ||||||| ||||
Db  97849 AATTGCGACTCTCAGAAGCATTTTTGGAACGTCTCTGGGAAAAAAAAATGCAGGATCCAA 97908

Qy   1249 gtaatcctgccatcatcaggcaggctgctggaaattatattggaagctttttggcaaagag 1308
          |||||||| ||||||||||  |||||||||||||||||||||||| |||||||| ||
Db  97909 ATAATCCTGTCATCATCAGGTGAGCTGCTGGAAATTATATTGGAAGCTCTTTGGCAAAAG 97968

Qy   1309 ctaaatttattcctcttatt---actgtaaaatcatgcctagatcttttgtttaactggc 1365
          ||||||| |||||||| |||   ||| ||||||||||||||| ||||| ||||||||||
Db  97969 CTAAATTCATTCCTCTCATTACAACTATAAAATCATGCCTAGGTCTTTCGGTTAACTGGC 98028

Qy   1366 tgcacatatacccttaataaccaggattcgggaacaaaggcattctgcgatgttgctctcc 1425
            |||| |||||||| |||||||||| | |||  |||| || || ||| |||||||||
Db  98029 CACACAAATACCTTACTAACCAGGATCTGAGAATGAAGGCTTTTTGTGATATTGCTCTCC 98088

Qy   1426 atggaccatttttactcagcctgccaagctgtgttctacaccttttgtttttagaacacaagc 1485
          ||  | |||||||||||||| |||||||||||| |||||||||||||||||||||||||
Db  98089 ATCAAACATTTTACTCAGCCTTCCAAGCTGTGTTCCACACCTTTGTTTTTAGACACAAGC 98148

Qy   1486 agcttttgagcggaaacctgaaagaaggtttgcagtatcttcagagtctgaattttgaagc 1545
          | ||||| |||| |||||||||||||||||    | ||||||| | ||  |||| |
Db  98149 AACTTTTAAGCGTAAACCTGAAAGAAGGTTTGAGATGTCTTCAGTGCCTACATTTTTAAG 98208

Qy   1546 ggataatgatgaagccagctaaatcccctgaaagatttgcctgccctcagtggttaacttttt 1605
          |||||||||  || |||||||||||| ||||| |||||||||||||||||||||||||| |
Db  98209 GGATAGTGACTAGGCAGCTAAATCCTCTGAATATTTGCCTGCCCTCAGTGGTTAACTTAT 98268

Qy   1606 ttgctgcaatcacaaataagtaccagctcgtcttctgctacaccatcattgaagaggaaca 1665
          |||||||  |||| ||||| || ||||| |   |||||||||| ||||||||||||||||
Db  98269 TTGCTGCTGTCACGAATAAATATCAGCTAG---TCTGCTACACTATCATTGAGAGGAACA 98325

Qy   1666 atcgccagatgctgccagtcattaggagtaccgctggaggagactcagtgcagatctgca 1725
          |  |||||                | ||| ||||||||||||| ||| |  |||| |||
Db  98326 ACGGCCAGA----------------AATACAGCTGGAGGAGACTCGGTGTGAACGTGCA 98368

Qy   1725 caaacccgctggacaccttcttcccctttgatccctgtgtgctgaaagaggtcaaagaaat 1785
          | |||||  ||||| |||||||||||||||||| |||| ||| ||| |||||| ||||
Db  98369 CCAACCCATTGGACGCCTTCTTCCCCTTTGATCCTTGTGCGCTTAAGCAGTCAAGGAAAG 98428

Qy   1786 tcattgatcctatttatcaggtatgggaagacatgagtgctgaagagctacaggagttca 1845
          |||||||| |||||||||| |||||||||||| |||| ||||||| |||||||| |
Db  98429 TCATTGATTCTATTTATCAGATATGGGAAGACATAAGTGTCGAAGAGCTTCAGGAGTTTA 98488

Qy   1846 agaaacccatgaaaaaggacatagtggaagatgaaagatgatgacttttctgaaaggcgaag 1905
          ||||||| |||||||||| || |   |||||| |||||||    || ||||||| ||||
Db  98489 AGAAACCTATGAAAAAGGAGATGGAAAAAGATGGAGATGAT---CTTTTGAAAGGTGAAG 98545

Qy   1906 tgccccagaatgatacccgtgattgggatcacaccaagctccttttgacacg   catttt 1960
          ! ||||  |||||||||| |||| ||||| | ||| |||  ||||||  ||     ||||
Db  98546 TTCCCCCAAATGATACCATGATCGGGATTATACCGAGCCACTTTGATGCGATGCACATTT 98605

Qy   1961 ccgaaagtccttcaagtagtgtgggctcccccaccgtgttgtacatg-caaccccagtcccc 2019
          || |||||||||||||||||||||||||||||| |||||  || || ||  ||||| |
Db  98606 CCAAAGTCCTTCAAGTAGTGTGGGCTCCCCACCTGTGTTACACCTGCCAGACCAGTTCCT 98665

Qy   2020 tctgacggcaggaaatttgtgactgagatgtgacatt 2055
          ||| |   |    |||||||| |||||||   ||||
Db  98666 TCTCATCTCGAGGATTTGTGATTGAGATGCAGCATT 98701


RESULT   5
AC013444
LOCUS       AC013444               192815 bp    DNA     linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens BAC clone RP11-470B22 from 2, complete sequence.
ACCESSION   AC013444
VERSION     AC013444.8  GI:18093267
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 192815)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 192815)
  AUTHORS   Waliqorski,J., Haakenson,W. and Spalding,L.
  TITLE     The sequence of Homo sapiens BAC clone RP11-470B22
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 192815)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 192815)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jan 9, 2002 this sequence version replaced gi:13518264.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            ------- ------ Summary Statistics
                Center project name: H_NH0470B22
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
```

```
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 52:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is AC024616; the clone sequenced to
            the right is RP11 190J23.  Actual start of this clone is at base
            position 1 of RP11 470B22; actual end is at base position 192815 of
            RP11-470B22.

            Unresolved homopolymeric run from base position 164727 to 164753.
FEATURES             Location/Qualifiers
     source          1..192815
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-470B22"
                     /clone_lib="RPCI-11"
     repeat_region   1..575
                     /rpt_family="L1"
     repeat_region   611..878
                     /rpt_family="ERVL"
     repeat_region   879..1042
                     /rpt_family="(TAGA)n"
     repeat_region   1046..1092
                     /rpt_family="ERVL"
     repeat_region   1987..2446
                     /rpt_family="L2"
     repeat_region   2737..2972
                     /rpt_family="L1"
     repeat_region   3702..4739
                     /rpt_family="AT_rich"
     misc_feature    4870..4891
                     /note="match to EST BE071256 (NID:q8416143)"
     repeat_region   6880..6992
                     /rpt_family="MIR"
     repeat_region   8677..8934
                     /rpt_family="Alu"
     repeat_region   8773..8800
                     /rpt_family="AT_rich"
     repeat_region   8955..8989
                     /rpt_family="AT_rich"
     repeat_region   9207..9247
                     /rpt_family="L2"
     repeat_region   9342..9773
                     /rpt_family="L2"
     repeat_region   9977..10011
                     /rpt_family="(TTTA)n"
     repeat_region   10976..11265
                     /rpt_family="Alu"
     repeat_region   12178..12207
                     /rpt_family="(T)n"
     repeat_region   12362..12390
                     /rpt_family="AT_rich"
     repeat_region   12702..12731
                     /rpt_family="(TG)n"
     repeat_region   12731..12758
                     /rpt_family="(GA)n"
     repeat_region   12758..12785
                     /rpt_family="(TG)n"
     repeat_region   13033..13054
                     /rpt_family="(CAAAA)n"
     repeat_region   13195..13699
                     /rpt_family="L1"
     repeat_region   13543..13567
                     /rpt_family="AT_rich"
     repeat_region   14142..14305
                     /rpt_family="MIR"
     repeat_region   14500..14534
                     /rpt_family="AT_rich"
     repeat_region   14617..14735
                     /rpt_family="MIR"
     repeat_region   14736..14836
                     /rpt_family="MER2_type"
     repeat_region   14837..15106
                     /rpt_family="Alu"
     repeat_region   15107..15317
                     /rpt_family="MER2_type"
     repeat_region   15318..15385
                     /rpt_family="MIR"
     repeat_region   15578..16066
                     /rpt_family="ERV1"
     repeat_region   16072..16112
                     /rpt_family="AT_rich"
     repeat_region   18192..18222
                     /rpt_family="AT_rich"
     repeat_region   18612..18899
                     /rpt_family="Alu"
     misc_feature    19181..19485
                     /note="match to EST R89091 (NID:q953918) yq02104.s1"
     misc_feature    19879..21714
                     /note="similar to Rattus norvegicus EST Y16973
                     (NID:q4727035)"
     misc_feature    19890..20284
                     /note="match to EST R89197 (NID:q954024) yq02104.r1"
     misc_feature    19915..20369
                     /note="similar to Homo sapiens EST BG545886
                     (NID:q13527431)"
     misc_feature    19924..20378
                     /note="similar to Homo sapiens EST AA213789 (NID:q1812416)
                     zr91a11.s1"
     misc_feature    19924..20369
                     /note="similar to Homo sapiens EST BG428305
                     (NID:q13334811)"
     misc_feature    20091..20478
                     /note="similar to Homo sapiens EST A1742153 (NID:q5110441)
                     wq49t01.x1"
     misc_feature    20158..20664
                     /note="similar to Homo sapiens EST AV703279
                     (NID:q10720608)"
     misc_feature    20243..21062
                     /note="similar to Mus musculus EST BG295610
                     (NID:q13057417)"
     misc_feature    20412..21044
                     /note="similar to Homo sapiens EST BE734354
                     (NID:q10148346)"
     misc_feature    20567..20666
                     /note="similar to Homo sapiens EST BG778634
                     (NID:q14048951)"
     misc_feature    20592..20963
                     /note="similar to Homo sapiens EST AA481295 (NID:q2210847)
                     zv02e05.r1"
     misc_feature    20703..21362
                     /note="similar to Homo sapiens EST BG502871
                     (NID:q13464388)"
     misc_feature    20798..21056
```

```
                     /note="similar to Homo sapiens EST AV649508
                     (NID:g9870522)"
     misc_feature    20862..21269
                     /note="similar to Homo sapiens EST AA191111 (NID:g1779805)
                     zp86a03.r1"
     misc_feature    20930..21593
                     /note="similar to Homo sapiens EST BF203684
                     (NID:g11097270)"
     misc_feature    20930..21038

  Query Match              58.6%;  Score 1211.6;  DB 9;  Length 192815;
  Best Local Similarity    80.2%;  Pred. No. 0;
  Matches 1617;  Conservative    0;  Mismatches  334;  Indels   65;  Gaps   14;

Qy         60 ttcggcccaatggcggcaccgctgcttcacacgcgtttgccgggagatgcggccgcttcg 119
Db      19803 TTAAGCCCCGGGGCAACCAGGCTGCTTCTCACATGGTTGCCGGG-GAAGCAGAGGCTTCA 19861

Qy        120 tcctctgcagttaagaagctgggcgcgtcgaggactgggatttcaaatatgcgtgcatta 179
Db      19862 GCTTCTGTGGTCAAGAAGCTGGGTATGTCATGGACCAGTATTTCAGATATGCTTGCATTC 19921

Qy        180 gagaatgacttttttcaattctcccccaagaaaaactgttcggtttggtggaactgtgaca 239
Db      19922 AAGAATGATTTTTCCAGCTCTCCCCCAAGAAAAATTGTTCATCTTGATGGAACGGGGACA 19981

Qy        240 gaagtcttgctgaagtacaaaaagggtgaaacaaatgactttgagttgttgaagaaccag 299
Db      19982 GAAGTCTTGCCGAAGGACAAAGAGTGAAACAAATTTAACTTTGAGCTGTTGAAGAGCCAG 20041

Qy        300 ctgttagatccagacataaaggatgaccagatcatcaactggctgctagaattccgttct 359
Db      20042 TTGTGAGATTCAAACATAAAGGATGACCAGATCTTGAACTGGCTGTGAAAATTTTGCTCT 20101

Qy        360 tctatcatgtacttgacaaaagactttgagcaacttatcagtattatattaagattgcct 419
Db      20102 TCTATCATGTACTTGGCAAAAGACTTTGAGCAGCTCCTCAGTATTACATTGAGATTGCCT 20161

Qy        420 tggttgaatagaaagtcaaacagtagtggaagagtatttggcttttcttggtaatcttgta 479
Db      20162 TGGTTGAATAGAATTCAAACAGTGGTGGAAGAGTATTTGGCTTCTCTGGGTAGTCTTGTA 20221

Qy        480 tcagcacagactgttttcctcagaccgtgtctcagcatgattgcttcccattttgtg--- 536
Db      20222 ACAACAGA----CTCTTCCTTAGCCCAAGTCTCAGCATAATTGCTTCCCATTTGGCGTAC 20277

Qy        537 cctccccgagtgatcattaaggaaggcgatgtagatgtttcagattctgatgatgaagat 596
Db      20278 CCTCCTTGAATGATCATTAAGAAGAGCCATGTAAATGTTTCAGATTCTGATGATGGAGAT 20337

Qy        597 gataatcttcctgcaaattttgacacatgtcacagagccttgcaaataatagcaagatat 656
Db      20338 GATTATCTTCCTGCAGATTTTGACACACGTCATACAACCTTACAAATAACAGCAAGATAT 20397

Qy        657 gtacccatcgacaccgtggtttctcatgccaatactggtggaaaaatttccatttgttcga 716
Db      20398 GTCCCACTGGCACTGTGGTTTCTTTTGCCCATACTGGTGGAAACATTTCCATTTGTTC-- 20455

Qy        717 aaatcagagagaacactggaatgttacgttcataacttactaaggattagtgtatatttt 776
Db      20456 -AATCAGAGAGAACACTGGAATGTTACATTCATAACTTACTAAAGATTAGTATATACTTT 20514

Qy        777 ccaaccttgaggcatgaaattctggagcttattattgaaaaactactcaagttggatgtg 836
Db      20515 CCAACCTTGAGGAATGAAATTCTGAAGCTTATTATTAAAAATCTATTCAAATTGGATGTG 20574

Qy        837 aatgcatcccggcagggtattgaagatgctgaagaaacagcaactcaaacttgtggtggg 896
Db      20575 AATACATTCTGGCATGATATTGAAGATGCTGAGGAAACAGCAGCTCAAACTTGTGGTGGG 20634

Qy        897 acagattccacggaaggattgtttaatatggatgaagatgaagaaactgaacatgaaaca 956
Db      20635 ACAGATTCCATGGAATGATTGTTCAATATGAACGAAGATGAAGAGAC------TGTTTCT 20688

Qy        957 aaggctggtcctgaacggctcgaccagatggtgcatcctgtagccgagcgcctggacatc 1016
Db      20689 AAGGCTGACTCCAAATGGCTCGGTCAGATGGTGCACCCTGTAGCTGAATGCCTGGACACC 20748

Qy       1017 ctgatgtctttggttttgtcctacatgaaggatgtctgctatg-----tagatggtaagg 1071
Db      20749 CTGCTGTCTTTGCTTTTGCAGTACTTGAAGTATGTCTGCTATGTAAAATACATGGTAAGG 20808

Qy       1072 ttgataacggcaaaacaaaggatctatatcgcgacctgataaacatctttgacaaactcc 1131
Db      20809 TTGATAACAACAAAATAAAGGATTTATA--------ATATATCACAAACTGATAAACATA 20860

Qy       1132 tgttgcccacccatgcctcctgccatgtacagttttttcatgttttacctctgtagtttca 1191
Db      20861 TTTGACAAACCCCTGCCTCCTGTCACGTACAGTTTTTTATGTTTTACCTCTGTGGCTTAA 20920

Qy       1192 aattgggattcgcagaggcattttggaacatctctgg---aaaaaattgcaggacccaa 1248
Db      20921 AATTGCGACTCTCAGAAGCATTTTTGGAACGTCTCTGGGAAAAAAAAATGCAGGATCCAA 20980

Qy       1249 gtaatcctgccatcatcaggcaggctgctggaaattatattggaagctttttggcaagag 1308
Db      20981 ATAATCCTGTCATCATCAGGTGAGCTGCTGGAAATTATATTGGAAGCTCTTTGGCAAAAG 21040

Qy       1309 ctaaatttattcctcttatt---actgtaaaatcatgcctagatcttttggttaactggc 1365
Db      21041 CTAAATTCATTCCTCTCATTACAACTATAAAATCATGCCTAGGTCTTTCGGTTAACTGGC 21100

Qy       1366 tgcacatatacсttaataaccaggattcgggaacaaaggcattctgcgatgttgctctcc 1425
Db      21101 CACACAAATACCTTACTAACCAGGATCTGAGAATGAAGGCTTTTTGTGATATTGCTCTCC 21160

Qy       1426 atggaccatttttactcagcctgccaagctgtgttctacacctttgtttttagacacaagc 1485
Db      21161 ATCAAACATTTTACTCAGCCTTCCAAGCTGTGTTCCACACCTTTGTTTTTAGACACAAGC 21220

Qy       1486 agctttgagcggaaacctgaaagaaggtttgcagtatcttcagagtctgaattttgagc 1545
Db      21221 AACTTTTAAGCGTAAACCTGAAAGAAGGTTTGAGATGTCTTCAGTGCCTACATTTTTAAG 21280

Qy       1546 ggatagtgatgagccagctaaatcccctgaagatttgcctgccctcagtggttaactttt 1605
Db      21281 GGATAGTGACTAGGCAGCTAAATCCTCTGAATATTTGCCTGCCCTCAGTGGTTAACTTAT 21340

Qy       1606 ttgctgcaatcacaaataagtaccagctcgtcttctgctacaccatcattgagaggaaca 1665
Db      21341 TTGCTGCTGTCACGAATAAATATCAGCTAG---TCTGCTACACTATCATTGAGAGGAACA 21397

Qy       1666 atcgccagatgctgccagtcattaggagtaccgctggaggagactcagtgcagatctgca 1725
Db      21398 ACGGCCAGA-----------------AATACAGCTGGAGGAGACTCGGTGTGAACGTGCA 21440

Qy       1726 caaacccgctggacaccttcttccccttttgatccctgtgtgctgaagaggtcaaagaaat 1785
Db      21441 CCAACCCATTGGACGCCTTCTTCCCCTTTGATCCTTGTGCGCTTAAGCAGTCAAGGAAAG 21500

Qy       1786 tcattgatcctatttatcaggtatgggaagacatgagtgctgaagagctacaggagttca 1845
Db      21501 TCATTGATTCTATTTATCAGATATGGGAAGACATAAGTGTCGAAGAGCTTCAGGAGTTTA 21560

Qy       1846 agaaacccatgaaaaaggacatagtggaagatgaagatgatgactttctgaaaggcgaag 1905
Db      21561 AGAAACCTATGAAAAAGGAGATGGAAAAAGATGGAGATGAT---CTTTTGAAAGGTGAAG 21617

Qy       1906 tgcccccagaatgataccgtgattgggatcacaccaagctcctttgacacg-----cattt 1960
Db      21618 TTCCCCCAAATGATACCATGATCGGGATTATACCGAGCCACTTTGATGCGATGCACATTT 21677

Qy       1961 ccgaagtccttcaagtagtgtgggctccccacccgtgttgtacatg-caacccagtcccc 2019
Db      21678 CCAAAGTCCTTCAAGTAGTGTGGGCTCCCCACCTGTGTTACACCTGCCAGACCAGTTCCT 21737
```

```
Qy   2020 tctgacggcagaaatttgtgactgagatgtgacatt 2055
          |||  |    |    ||||||| |||||||    ||||
Db  21738 TCTCATCTCGAGGATTTGTGATTGAGATGCAGCATT 21773


RESULT    6
AK055742
LOCUS       AK055742                2493 bp    mRNA    linear   PRI 31-OCT-2001
DEFINITION  Homo sapiens cDNA FLJ31180 fis, clone KIDNE2000266.
ACCESSION   AK055742
VERSION     AK055742.1  GI:16550546
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens kidney cDNA to mRNA, clone_lib:KIDNE2
            clone:KIDNE2000266.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,
            Ono,Y., Hotuta,T., Hiraoka,S., Murakawa,K., Takiguchi,S.,
            Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,
            Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,
            Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y.,
            Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y.,
            Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2493)
  AUTHORS   Isogai,T., Otsuki,T. and Sugiyama,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: RAB and
            HRI.
FEATURES             Location/Qualifiers
     source          1..2493
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="KIDNE2000266"
                     /tissue_type="kidney"
                     /clone_lib="KIDNE2"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      554 a    700 c    761 g    478 t
ORIGIN


  Query Match             27.4%;  Score 567;  DB 9;  Length 2493;
  Best Local Similarity   98.3%;  Pred. No. 5.7e-146;
  Matches  573;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy   1196 gggattcgcagaggcatttttggaacatctctggaaaaaattgcaggacccaagtaatcc 1255
          |||||||||||||||||||||||||||||| ||||||||||||||||| |||||||||||
Db    711 GGGATTCGCAGAGGCATTTTTGGAACATCTTTGGAAAAAATTGCAGGATCCAAGTAATCC 770

Qy   1256 tgccatcatcaggcaggctgctggaaattatattggaagctttttggcaagagctaaatt 1315
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    771 TGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAGAGCTAAATT 830

Qy   1316 tattcctcttattactgtaaaatcatgcctagatcttttggttaactggctgcacatata 1375
          |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db    831 TATTCCTCTTATTACTGTAAAACCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATA 890

Qy   1376 ccttaataaccaggattcgggaacaaaggcattctgcgatgttgctctccatggaccatt 1435
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    891 CCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCAATGTTGCTCTCCATGGACCATT 950

Qy   1436 ttactcagcctgccaagctgtgttctacacctttgtttttagacacaagcagcttttgag 1495
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    951 TTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGAG 1010

Qy   1496 cggaaacctgaaagaaggtttgcagtatcttcagagtctgaattttgagcggatagtgat 1555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1011 CGGAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGAT 1070

Qy   1556 gagccagctaaatcccctgaagatttgcctgccctcagtggttaactttttgctgcaat 1615
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1071 GAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGCAAT 1130

Qy   1616 cacaaataagtaccagctcgtcttctgctacaccatcattgagaggaacaatcgccagat 1675
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   1131 CACAAATAAGTACCAGCTCGTCTTCTGCTACACCATCATCGAGAGGAACAATCGCCAGAT 1190

Qy   1676 gctgccagtcattaggagtaccgctggaggagactcagtgcagatctgcacaaacccgct 1735
          |||||||||||||||||||||| ||||||||||||||||||||| |||||||||||| ||
Db   1191 GCTGCCAGTCATTAGGAGTACCACTGGAGGAGACTCAGTGCAGACCTGCACAAACCCACT 1250

Qy   1736 ggacaccttcttcccctttgatccctgtgtgctgaagaggtca 1778
          |||||||||| ||||||||||||||||||||||||||||||| |
Db   1251 GGACACCTTCGTCCCCTTTGATCCCTGTGTGCTGAAGAGGTGA 1293


RESULT    7
BC009198
LOCUS       BC009198                1177 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, Similar to RNA polymerase I transcription factor
            RRN3, clone MGC:15321 IMAGE:3678732, mRNA, complete cds.
ACCESSION   BC009198
VERSION     BC009198.1  GI:14327947
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1177)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 24 Row: k Column: 14
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 3355302.
FEATURES             Location/Qualifiers
     source          1..1177
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:15321 IMAGE:3678732"
                     /tissue_type="Lymph, Burkitt lymphoma"
                     /clone_lib="NIH_MGC_8"
                     /lab_host="DH10B R"
                     /note="Vector: pOTB7"
     CDS             316..636
```

```
                     /codon_start=1
                     /product="Similar to RNA polymerase I transcription factor
                     RRN3"
                     /protein_id="AAH09198.1"
                     /db_xref="GI:14327948"
                     /translation="MRALENDFFNSPPRKTVRFGGTVTEVLLKYKKGETNDFELLKNQ
                     LLDPDIKDDQIINWLLEFRSSVMYLTKDFEQLISIILRLPWLNRSQTVVEEYLAFLGN
                     LVSA"
BASE COUNT      339 a    222 c    253 g    363 t
ORIGIN

  Query Match             24.5%;  Score 506.4;  DB 9;  Length 1177;
  Best Local Similarity   88.5%;  Pred. No. 3.3e-129;
  Matches  549;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        157 ggatttcaaatatgcgtgcattagagaatgactttttcaattctcccccaagaaaaactg 216
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        305 GGATTTCAAATATGCGTGCATTAGAGAATGATTTTTTCAATTCTCCCCCAAGAAAAACTG 364

Qy        217 ttcggtttggtggaactgtgacagaagtcttgctgaagtacaaaaagggtgaaacaaatg 276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 TTCGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATG 424

Qy        277 actttgagttgttgaagaaccagctgttagatccagacataaaggatgaccagatcatca 336
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 ACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCA 484

Qy        337 actggctgctagaattccgttcttctatcatgtacttgacaaaagactttgagcaactta 396
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        485 ACTGGCTGCTAGAATTCCGTTCTTCTGTCATGTACTTGACAAAAGACTTTGAGCAACTTA 544

Qy        397 tcagtattatattaagattgccttggttgaatagaagtcaaacagtagtggaagagtatt 456
              ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        545 TCAGTATTATATTGAGATTGCCTTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATT 604

Qy        457 tggcttttcttggtaatcttgtatcagcacagactgttttcctcagaccgtgtctcagca 516
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        605 TGGCTTTTCTTGGTAATCTTGTATCAGCATAGACTGTTTTCCTCAGACCGTGTCTCAGCA 664

Qy        517 tgattgcttcccattttgtgcctccccgagtgatcattaaggaaggcgatgtagatgttt 576
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        665 TGATTGCTTCCCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTT 724

Qy        577 cagattctgatgatgaagatgataatcttcctgcaaattttgacacatgtcacagagcct 636
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db        725 CAGATTCTGATGATGAAGATGATAATCTTCCTGCAAATTTTGACACATATCACAGAGCCT 784

Qy        637 tgcaaataatagcaagatatgtaccatcgacaccgtggtttctcatgccaatactggtgg 696
              ||||||||||||||||||||||||||     |    |   |||   | |        | || |
Db        785 TGCAAATAATAGCAAGATATGTACCATCATGAGTATACTTTTCCTTATTTTGAATGTTTA 844

Qy        697 aaaaatttccatttgttcgaaaatcagagagaacactggaatgttacgttcataacttac 756
              |         |   |   |       ||| ||  | ||    |   |||||||      || |
Db        845 ATTCTCAAGAAAATTGTAATCAATTAGTAAAAATTATAAAATGTTAATAGTATTAAAGCT 904

Qy        757 taaggattagtgtatatttt 776
              | ||  |||    |  |||||
Db        905 TGAGTCTTACATTGCATTTT 924

RESULT    8
AX274907
LOCUS       AX274907                1461 bp    DNA     linear   PAT 29-OCT-2001
DEFINITION  Sequence 172 from Patent WO0172777.
ACCESSION   AX274907
VERSION     AX274907.1  GI:16547539
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Hillman,J.L., Baughn,M.R., Yue,H., Lal,P., Lu,D.A., Patterson,C.,
            Azimzai,Y., Bandman,O., Tang,Y.T., Mathur,P., Shah,P., Au Young,J.
            and Reddy,R.
  TITLE     Transcription factors
  JOURNAL   Patent: WO 0172777-A 172 04-OCT-2001;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..1461
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 2646274CB1"
BASE COUNT      413 a    335 c    298 g    415 t
ORIGIN

  Query Match             20.3%;  Score 420.8;  DB 6;  Length 1461;
  Best Local Similarity   97.3%;  Pred. No. 2e-105;
  Matches  428;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1188 ttcaaattgggattcgcagaggcatttttggaacatctctggaaaaaattgcaggaccca 1247
              || ||  ||||||||||||||||||||||||||||||| |||||||| |||||||| |||
Db        530 TTTAATATGGGATTCGCAGAGGCATTTTTGGAACATCTTTGGAAAAACTTGCAGGATCCA 589

Qy       1248 agtaatcctgccatcatcaggcaggctgctggaaattatattggaagctttttggcaaga 1307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        590 AGTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAGA 649

Qy       1308 gctaaatttattcctcttattactgtaaaatcatgcctagatcttttggttaactggctg 1367
              |||||||||||| ||||||||||||||||| |||||||||||||||||||||||||||||
Db        650 GCTAAATTTATTTCTCTTATTACTGTAAAACCATGCCTAGATCTTTTGGTTAACTGGCTG 709

Qy       1368 cacatataccttaataaccaggattcgggaacaaaggcattctgcgatgttgctctccat 1427
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        710 CACATATACCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCAT 769

Qy       1428 ggaccatttactcagcctgccaagctgtgttctacacctttgtttttagacacaagcag 1487
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        770 GGACCATTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAG 829

Qy       1488 cttttgagcggaaacctgaaagaaggtttgcagtatcttcagagtctgaattttgagcgg 1547
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db        830 CTTTTGAGCGGAAACCTGAAAGAAGGTTTGCAGTATCCTCAGAGTCTGAATTTTGAGCGG 889

Qy       1548 atagtgatgagccagctaaatcccctgaagatttgcctgccctcagtggttaactttttt 1607
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        890 ATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTT 949

Qy       1608 gctgcaatcacaaataagta 1627
              |||||||||||||| | | |
Db        950 GCTGCAATCACAAAGATGAA 969

RESULT    9
BC006441
LOCUS       BC006441                2627 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, Similar to RNA polymerase I transcription factor
            RRN3, clone MGC:13169 IMAGE:3351791, mRNA, complete cds.
ACCESSION   BC006441
VERSION     BC006441.1  GI:13623642
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2627)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
```

```
            Email: cqapbs t*mail.nih.qov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencinq by: Genome Sequence Centre,
            BC Cancer Aqency, Vancouver, BC, Canada
            info*beqsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Cham, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Linq, Carrie Mathewson, Candice McIeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
            Georqe Yanq, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            throuqh the I.M.A.G.E. Consortium/LLNL at: http://imaqe.llnl.qov
            Series: IRAL Plate: 18 Row: d Column: 24
            This clone was selected for full lenqth sequencinq because it
            passed the followinq selection criteria: Hexamer frequency ORF
            analysis.
FEATURES             Location/Qualifiers
     source          1. .2627
                     /orqanism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "MGC:14169 IMAGE:4351791"
                     /tissue_type "Eye, retinoblastoma"
                     /clone_lib "NIH_MGC_16"
                     /lab_host "DH10B_R"
                     /note "Vector: pOTB7"
     CDS             1698. .2156
                     /codon_start 1
                     /product "Similar to RNA polymerase I transcription factor
                     RRN4"
                     /protein_id "AAH06441.1"
                     /db_xref "GI:13623643"
                     /translation "MGFAEAFLEHLWKNLQDPSNPAIIRQAAGNYIGSFLARAKFISL
                     ITVKPCLDLLVNWLHIYLNNQDSGTKAFCDVALHGPFYSACQAVFYTFVFRHKQLLSG
                     NLKEGLQYPQSLNFERIVMSQLNPLKICLPSVVNFFAAITKMKTCGYGWW"
BASE COUNT       837 a    492 c    525 q    773 t
ORIGIN


  Query Match             20.3%;  Score 420.8;  DB 9;  Length 2627;
  Best Local Similarity   97.3%;  Pred. No. 2.2e-105;
  Matches  428;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy    1188 ttcaaattqqqattcqcaqaqqcatttttqqaacatctctqqaaaaaattqcaqqaccca 1247
           || ||  ||||||||||||||||||||||||||||||| ||||||| ||||||| |||
Db    1692 TTTAATATGGGATTCGCAGAGGCATTTTTGGAACATCTTTGGAAAAACTTGCAGGATCCA 1751

Qy    1248 aqtaatcctqccatcatcaqqcaqqctqctqqaaattatattqqaaqctttttqqccaaqa 1307
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1752 AGTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAGA 1811

Qy    1308 qctaaatttatttcctcttattactqtaaaatcatqcctaqatctttqqttaactqqctq 1367
           |||||||||||| |||||||||||||||| ||||||||||||||||||||||||||||||
Db    1812 GCTAAATTTATTTCTCTTATTACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTG 1871

Qy    1368 catataccttaataaccaqqattcqqqaacaaaqqcattctqcqatqttqctctccat 1427
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1872 CATATATCCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCAT 1931

Qy    1428 qqaccatttactcaqcctqccaaqctqtattctacacctttqttttqqacacaaqcaq 1487
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1932 GGACCATTTACTCAGCCTGCCAAGCTGTATTCTACACTTTGTTTTTAGACACAAGCAG 1991

Qy    1488 ctttqaqcqqaaacctqaaqaaqqtttqcaqtatcttcaqaqtctqaattttqaqcqq 1547
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    1992 CTTTGAGCGGAAACCTGAAAGAAGGTTTGCAGTATCCTCAGAGTCTGAATTTTGAGCGG 2051

Qy    1548 ataqtqatqaqccaqctaaatcccctqaaaqatttqcctqcccteaqtqqttaacttttt 1607
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2052 ATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTT 2111

Qy    1608 qctqcaatcacaaataaqta 1627
           ||||||||||||||| | | |
Db    2112 GCTGCAATCACAAAGATGAA 2131


RESULT  10
AC040158
LOCUS       AC040158               129473 bp    DNA     linear   HTG 03-SEP-2000
DEFINITION  Homo sapiens chromosome 16 clone CTA-13384, WORKING DRAFT SEQUENCE,
            14 unordered pieces.
ACCESSION   AC040158
VERSION     AC040158.3  GI:9965546
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 129473)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 129473)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Sep 1, 2000 this sequence version replaced gi:9090652.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jqi.doe.qov
            -------
            Project Information
            Center Project Name: 0
            Center clone name: CTP978SKA_13384
            -------
            Summary Statistics
            Consensus quality: 90046 bases at least Q40
            Consensus quality: 99266 bases at least Q30
            Consensus quality: 104098 bases at least Q20
            Estimated insert size: 125000; aqarose-fp estimation
            Estimated insert size: 128173; sum-of-contiqs estimation
            Quality coverage:  4.51 in Q20 bases; aqarose-fp estimation
            Quality coverage:  4.4 in Q20 bases; sum-of-contiqs estimation.
            * NOTE: This is a 'workinq draft' sequence. It currently
            * consists of 14 contiqs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contiqs are represented as
            * runs of N, but the exact sizes of the qaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1      1718: contiq of 1718 bp in lenqth
            *      1719      1818: qap of unknown lenqth
            *      1819      3842: contiq of 2014 bp in lenqth
            *      3843      3942: qap of unknown lenqth
            *      3943      5567: contiq of 1635 bp in lenqth
            *      5568      5667: qap of unknown lenqth
            *      5668      7520: contiq of 1853 bp in lenqth
            *      7521      7620: qap of unknown lenqth
            *      7621      9146: contiq of 1526 bp in lenqth
            *      9147      9246: qap of unknown lenqth
            *      9247     11144: contiq of 1898 bp in lenqth
            *     11145     11244: qap of unknown lenqth
            *     11245     12746: contiq of 1502 bp in lenqth
            *     12747     12846: qap of unknown lenqth
            *     12847     14568: contiq of 1722 bp in lenqth
            *     14569     14668: qap of unknown lenqth
            *     14669     19396: contiq of 4728 bp in lenqth
            *     19397     19496: qap of unknown lenqth
```

```
            *    19497    25634: contig of 6138 bp in length
            *    25635    25734: gap of unknown length
            *    25735    36780: contig of 11046 bp in length
            *    36781    36880: gap of unknown length
            *    36881    50223: contig of 13343 bp in length
            *    50224    50323: gap of unknown length
            *    50324    71312: contig of 20989 bp in length
            *    71313    71412: gap of unknown length
            *    71413   129473: contig of 58061 bp in length.
FEATURES             Location/Qualifiers
     source          1. .129473
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="CTA-133B4"
                     /clone_lib="CalTech human BAC library A"
BASE COUNT    30715 a   30506 c   31241 g   35543 t    1468 others
ORIGIN


  Query Match             10.1%;  Score 208;  DB 2;  Length 129473;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-46;
  Matches  208;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1861 aqgacataqtqqaaqatqaaqatqatqacttttctqaaaqqcqaaqtqccccaqaatqata 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  123987 AGGACATAGTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCAGAATGATA 124046

Qy    1921 ccgtgattgggatcacaccaaqctcctttqacacgcatttccgaaqtccttcaaqtagtg 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124047 CCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAGTCCTTCAAGTAGTG 124106

Qy    1981 tqqqctcccqacccqtqttqtacatqcaacccaqtcccctctqacqqcaqaaatttqtqa 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124107 TGGGCTCCCCACCCGTGTTGTACATGCAACCCAGTCCCCTCTGACGGCAGAAATTTGTGA 124166

Qy    2041 ctqaqatqtqacatttqqqattccccat 2068
           |||||||||||||||||||||||||||||
Db  124167 CTGAGATGTGACATTTGGGATTCCCCAT 124194


RESULT  11
AC092137/c
LOCUS       AC092137               143900 bp    DNA     linear   HTG 22-JUN-2001
DEFINITION  Homo sapiens chromosome 16 clone RP11-344H15, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.
ACCESSION   AC092137
VERSION     AC092137.1  GI:14522983
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 143900)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 143900)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 538644
            Center clone name: RPCI-11_344H15
            -------
            Summary Statistics
            Consensus quality: 124071 bases at least Q40
            Consensus quality: 136140 bases at least Q40
            Consensus quality: 137741 bases at least Q20
            Estimated insert size: 211000; agarose-fp estimation
            Estimated insert size: 142200; sum-of-contigs estimation
            Quality coverage:  4.0 in Q20 bases; agarose-fp estimation
            Quality coverage:  5.94 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1033: contig of 1033 bp in length
            *     1034     1133: gap of unknown length
            *     1134     2547: contig of 1414 bp in length
            *     2548     2647: gap of unknown length
            *     2648     4205: contig of 1558 bp in length
            *     4206     4305: gap of unknown length
            *     4306     6564: contig of 2259 bp in length
            *     6565     6664: gap of unknown length
            *     6665     9065: contig of 2401 bp in length
            *     9066     9165: gap of unknown length
            *     9166    12059: contig of 2894 bp in length
            *    12060    12159: gap of unknown length
            *    12160    19295: contig of 7136 bp in length
            *    19296    19395: gap of unknown length
            *    19396    25668: contig of 6273 bp in length
            *    25669    25768: gap of unknown length
            *    25769    31013: contig of 5245 bp in length
            *    31014    31113: gap of unknown length
            *    31114    36555: contig of 5442 bp in length
            *    36556    36655: gap of unknown length
            *    36656    45338: contig of 8683 bp in length
            *    45339    45438: gap of unknown length
            *    45439    52918: contig of 7480 bp in length
            *    52919    53018: gap of unknown length
            *    53019    65720: contig of 12702 bp in length
            *    65721    65820: gap of unknown length
            *    65821    76109: contig of 10289 bp in length
            *    76110    76209: gap of unknown length
            *    76210    85439: contig of 9230 bp in length
            *    85440    85539: gap of unknown length
            *    85540   101620: contig of 16081 bp in length
            *   101621   101720: gap of unknown length
            *   101721   122163: contig of 20443 bp in length
            *   122164   122263: gap of unknown length
            *   122264   143900: contig of 21637 bp in length.
FEATURES             Location/Qualifiers
     source          1. .143900
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-344H15"
                     /clone_lib "RPCI human BAC library 11"
BASE COUNT    37791 a   34128 c   34320 g   35886 t    1775 others
ORIGIN


  Query Match             10.1%;  Score 208;  DB 2;  Length 143900;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-46;
  Matches  208;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1861 aqgacataqtqqaaqatqaaqatqatqacttttctqaaaqqcqaaqtqccccaqaatqata 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27338 AGGACATAGTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCAGAATGATA 27279

Qy    1921 ccgtgattgggatcacaccaaqctcctttqacacgcatttccgaaqtcctcaaqtagtg 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27278 CCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAGTCCTTCAAGTAGTG 27219

Qy    1981 tqqqctcccqacccqtqttqtacatqcaacccaqtcccctctqacqqcaqaaatttqtqa 2040
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27218 TGGGCTCCCCACCCGTGTTGTACATGCAACCCAGTCCCCTCTGACGGCAGAAATTTGTGA 27159

Qy   2041 ctgaqatqtqacatttqaqattccccat 2068
          |||||||||||||||||||||||||||||
Db  27158 CTGACATGTGACATTTGGGATTCCCCAT 27131


RESULT  12
AC007615
LOCUS       AC007615               175691 bp    DNA     linear   HTG 03-JUL-2001
DEFINITION  Homo sapiens chromosome 16 clone RP11-528K16, WORKING DRAFT
            SEQUENCE, 1 ordered pieces.
ACCESSION   AC007615
VERSION     AC007615.6  GI:14589428
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175691)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175691)
  AUTHORS   Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
            Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
            Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
            Campbell,C., Fawcett,J., Matthie,M., Bussod,M., Sutherland,R.,
            McMurry,K., Ban,C. and Deaven,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (20 MAY 1999) Center for Human Genome Studies, DOE Joint
            Genome Institute, Los Alamos National Laboratory, MS M888, Los
            Alamos, NM 87545, USA
COMMENT     On Jul 3, 2001 this sequence version replaced gi:13928651.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
                          .
                          .
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
                          .
                          .
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    175691: contig of 175691 bp in length.
FEATURES             Location/Qualifiers
     source          1..175691
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-528K16"
BASE COUNT    42092 a   42125 c   45277 g   45896 t       1 others
ORIGIN

          .         .


  Query Match              9.3%;  Score 192;  DB 2;  Length 175691;
  Best Local Similarity   89.2%;  Pred. No. 1.6e-41;
  Matches  207;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy   1293 ttatattqqaaqctttttqqcaaqaqctaaatttattcctcttattactqtaaaatcatq 1352
          || | |  |||| |  |  |  |     |  ||| || |   || ||||||||||||||||
Db  30961 TTTTTTAACAAGTGTAAAGTTGATTTGAATTTTTCTTTCCTCTAGTACTGTAAAATCATG 31020

Qy   1353 cctaqatcttttqattaactqqctqcacatataccttaataaccaqqattcqqqaacaaa 1402
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31021 CCTAGATCTTTTGTTAACTGGCTGCACATATACCTTAATAACCAGGATTCGGGAACAAA 31080

Qy   1403 qqcattctqccatattqctctccatqqaccatttactcaqcctqccaaqctqtqtctta 1462
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31081 GGCATTCTGCCATGTTGCTCTCCATGGACCATTTIACTCAGCCTGCCAAGCTGTGTCTA 31140

Qy   1463 cacctttatttttqaqacacaaqcaqctttqaqcqqaaacctqaaaqaaqt 1514
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31141 CACCTTTGTTTTTAGACACAAGCAGCTTTGAGCGGAAACCTGAAAGAAGT 31192


RESULT  13
AC009130
LOCUS       AC009130               191496 bp    DNA     linear   HTG 21-JUN-2000
DEFINITION  Homo sapiens chromosome 16 clone RP11-501P17, WORKING DRAFT
            SEQUENCE, 43 unordered pieces.
ACCESSION   AC009130
VERSION     AC009130.6  GI:8575965
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191496)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191496)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jun 21, 2000 this sequence version replaced gi:7689948.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
              .  .
            Project Information
            Center Project Name: 599126
            Center clone name: RPCI 11_501P17

            Summary Statistics
            Consensus quality: 158439 bases at least Q40
            Consensus quality: 176841 bases at least Q30
            Consensus quality: 180633 bases at least Q20
            Estimated insert size: 201840; agarose fp estimation
            Estimated insert size: 187296; sum-of-contigs estimation
            Quality coverage:  3.4 in Q20 bases; agarose fp estimation
            Quality coverage:  3.67 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 43 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1071: contig of 1071 bp in length
            *    1072     1171: gap of unknown length
            *    1172     2358: contig of 1187 bp in length
```

```
*      2359      2458: gap of unknown length
*      2459      3680: contig of 1222 bp in length
*      3681      3780: gap of unknown length
*      3781      5278: contig of 1498 bp in length
*      5279      5378: gap of unknown length
*      5379      6522: contig of 1144 bp in length
*      6523      6622: gap of unknown length
*      6623      8589: contig of 1967 bp in length
*      8590      8689: gap of unknown length
*      8690     10266: contig of 1577 bp in length
*     10267     10366: gap of unknown length
*     10367     11714: contig of 1348 bp in length
*     11715     11814: gap of unknown length
*     11815     13443: contig of 1629 bp in length
*     13444     13543: gap of unknown length
*     13544     14645: contig of 1102 bp in length
*     14646     14745: gap of unknown length
*     14746     15751: contig of 1006 bp in length
*     15752     15851: gap of unknown length
*     15852     16868: contig of 1017 bp in length
*     16869     16968: gap of unknown length
*     16969     19037: contig of 2069 bp in length
*     19038     19137: gap of unknown length
*     19138     20794: contig of 1657 bp in length
*     20795     20894: gap of unknown length
*     20895     22876: contig of 1982 bp in length
*     22877     22976: gap of unknown length
*     22977     24751: contig of 1775 bp in length
*     24752     24851: gap of unknown length
*     24852     27381: contig of 2530 bp in length
*     27382     27481: gap of unknown length
*     27482     30339: contig of 2858 bp in length
*     30340     30439: gap of unknown length
*     30440     32397: contig of 1958 bp in length
*     32398     32497: gap of unknown length
*     32498     34880: contig of 2383 bp in length
*     34881     34980: gap of unknown length
*     34981     37578: contig of 2598 bp in length
*     37579     37678: gap of unknown length
*     37679     40004: contig of 2326 bp in length
*     40005     40104: gap of unknown length
*     40105     43544: contig of 3440 bp in length
*     43545     43644: gap of unknown length
*     43645     46429: contig of 2785 bp in length
*     46430     46529: gap of unknown length
*     46530     49686: contig of 3157 bp in length
*     49687     49786: gap of unknown length
*     49787     52269: contig of 2483 bp in length
*     52270     52369: gap of unknown length
*     52370     55195: contig of 2826 bp in length
*     55196     55295: gap of unknown length
*     55296     58516: contig of 3221 bp in length
*     58517     58616: gap of unknown length
*     58617     61359: contig of 2743 bp in length
*     61360     61459: gap of unknown length
*     61460     66151: contig of 4692 bp in length
*     66152     66251: gap of unknown length
*     66252     73145: contig of 6894 bp in length
*     73146     73245: gap of unknown length
*     73246     78200: contig of 4955 bp in length
*     78201     78300: gap of unknown length
*     78301     83186: contig of 4886 bp in length
*     83187     83286: gap of unknown length
*     83287     88570: contig of 5284 bp in length
*     88571     88670: gap of unknown length
*     88671     95276: contig of 6606 bp in length
*     95277     95376: gap of unknown length
*     95377    102234: contig of 6858 bp in length
*    102235    102334: gap of unknown length
*    102335    107630: contig of 5296 bp in length
*    107631    107730: gap of unknown length
*    107731    120220: contig of 12490 bp in length
*    120221    120320: gap of unknown length
*    120321    134451: contig of 14131 bp in length
*    134452    134551: gap of unknown length
*    134552    144891: contig of 10340 bp in length
*    144892    144991: gap of unknown length
*    144992    159434: contig of 14443 bp in length
*    159435    159534: gap of unknown length
*    159535    175863: contig of 16329 bp in length
*    175864    175963: gap of unknown length
*    175964    191496: contig of 15533 bp in length.
FEATURES             Location/Qualifiers
     source          1..191496
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-501P17"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT    47031 a  47510 c  47005 g  45732 t   4218 others
ORIGIN


  Query:Match              9.3%;  Score 192;  DB 2;  Length 191496;
  Best Local Similarity   89.2%;  Prob. No. 1.7e-41;
  Matches  207;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy    1283 ttatattgaaagctttttgagcaagagctaaatttattcctcttattactgtaaaatcatg 1342
           || | |  ||||  |   |  |     |  ||| || |   || ||||||||||||||||
Db   26578 TTTTTTATGAAGTGTAAAGTTGATTTGAATTTTTCTTTCCTCTAGTACTGTAAAATCATG 26637

Qy    1343 cctagatcttttgtttaactggctgcacatatacctttaataaccaggattcgggaacaaa 1402
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   26638 CCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCAGGATTCGGGAACAAA 26697

Qy    1403 ggcattctgcgatgttgctctccatggaccatttactcaggctgccaaggctgtgttcta 1462
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   26698 GGCATTCTGCGATGTTGCTCTCCATGGACCATTTACTCAGCCTGCCAAGCTGTGTTCTA 26757

Qy    1463 cacctttgttttagacacaagcagctttgagcggaaacctgaaagaaggt 1514
           ||||||||||:|||||||||||||||||||||||||||||||||||||||
Db   26758 CACCTTTGTTTTAGACACAAGCAGCTTTGAGCGGAAACCTGAAAGAAGGT 26809


RESULT  14
AC009093
LOCUS       AC009093              208008 bp    DNA     linear   HTG 04 SEP 2001
DEFINITION  Homo sapiens chromosome 16 clone RP11-426C22, WORKING DRAFT
            SEQUENCE, 1 ordered pieces.
ACCESSION   AC009093
VERSION     AC009093.8  GI:15426058
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 208008)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 208008)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Sep 4, 2001 this sequence version replaced gi:13786300.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
            ----------
```

```
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base by base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
                          -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1   208008: contig of 208008 bp in length.
FEATURES             Location/Qualifiers
     source          1..208008
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "16"
                     /clone "RP11 426C22"
BASE COUNT    49489 a  50452 c  54814 g  54252 t       1 others
ORIGIN


  Query Match             9.4%;  Score 192;  DB 2;  Length 208008;
  Best Local Similarity   89.2%;  Pred. No. 1.7e-41;
  Matches  207;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy    1284 ttatattqaaaqctttttqqcaaqaqctaaatttattcctcttattactqtaaaatcatq 1342
           || | |  ||||  |   |   |      |  ||| || |   || ||||||||||||||||
Db   28443 TTTTTTATGAAGTGTAAAGTTGATTTGAATTTTTCTTTCCTCTAGTACTGTAAAATCATG 28502

Qy    1343 cctaqatcttttqqttaactqqctqcacatataccttaataaccaqqattcqqqaacaaa 1402
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   28503 CCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCAGGATTCGGGAACAAA 28562

Qy    1403 qqcattctcqcqatqttqctctccatqqaccatttactcaqcctqccaaqctqtqttcta 1462
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   28563 GGCATTCTGCGATGTTGCTCTCCATGGACCATTTACTCAGCCTGCCAAGCTGTGTTCTA 28622

Qy    1463 cacctttqttttaqacacaaqcaqctttqaqcqqaaacctqaaaqaaqqt 1514
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   28623 CACCTTTGTTTTTAGACACAAGCAGCTTTGAGCGGAAACCTGAAAGAAGGT 28674


RESULT  15
AC092562
LOCUS       AC092562              187643 bp    DNA     linear   HTG 18-OCT-2001
DEFINITION  Papio hamadryas clone RP41-285I13, WORKING DRAFT SEQUENCE, 8
            unordered pieces.
ACCESSION   AC092562
VERSION     AC092562.3  GI:16258968
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      baboon.
  ORGANISM  Papio hamadryas
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 187643)
  AUTHORS   Muzny,D.M., Adams,C., Adio-oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,B.,
            Duqan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraquto,D., Flaqq,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.R., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodqson,A., Hoques,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,D., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,K.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Mattindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Moi,G., Metzker,M., Miner,G., Miner,Z., Mitchell,I., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oraqunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Soderqren,E., Sonaike,T., Sparks,A., Stanley,R.,
            Stone,R., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washinaton,C.,
            Watlinqton,S., Williams,G., Williamson,A., Wlodzyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 187643)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, one
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 18, 2001 this sequence version replaced gi:15808491.
            ------------ Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
                ------- Project Information
                Center project name: ZOAG
                Center clone name: RP41-285I13
                ------ Summary Statistics
                Sequencing vector: Plasmid; M77789
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 191459 bases at least Q40
                Consensus quality: 193431 bases at least Q30
                Consensus quality: 194801 bases at least Q20
                Estimated insert size: 188848; sum of contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 8.3x in Q20 bases; sum-of contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    67581: contig of 67581 bp in length
            *   67582    67681: gap of unknown length
            *   67682   113789: contig of 46108 bp in length
            *  113790   113889: gap of unknown length
```

```
            *   113890   131405: contig of 17516 bp in length
            *   131406   131505: gap of unknown length
            *   131506   144468: contig of 12963 bp in length
            *   144469   144568: gap of unknown length
            *   144569   163180: contig of 18612 bp in length
            *   163181   163280: gap of unknown length
            *   163281   176573: contig of 13293 bp in length
            *   176574   176673: gap of unknown length
            *   176674   185462: contig of 8789 bp in length
            *   185463   185562: gap of unknown length
            *   185563   187643: contig of 2081 bp in length.
FEATURES             Location/Qualifiers
     source          1. .187643
                     /organism="Papio hamadryas"
                     /db_xref="taxon:9557"
                     /clone="RP41-285113"
BASE COUNT    49173 a  41946 c  44647 g  51158 t    719 others
ORIGIN


  Query Match             9.2%;  Score 190.8;  DB 2;  Length 187643;
  Best Local Similarity   94.3%;  Pred. No. 3.6e-41;
  Matches  198;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy    1859 aaaqqacatagtqqaaqatqaaqatgatgactttctqaaaggcqaaqtqccccaqaatqa 1918
           | |||| |||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   40557 ACAGGAGATAGTGGAAGATGAAGATGATGACTTTCTGAAAGGTGAAGTGCCCCAGAATGA 40616

Qy    1919 taccqtgattgggatcacaccaaqctcctttqacacqcatttccqaaqtccttcaaqtaq 1978
           ||||||||||||||||||| |||||||| ||  | ||||||||||||||||||||||||
Db   40617 TACCGTGATTGGGATCACACCGAGCTCCTTCGATGCACATTTCCGAAGTCCTTCAAGTAG 40676

Qy    1979 tgtgqgctccccaccgtgttgtacatgcaacccaqtcccctctqacqqcaqaaatttgt 2038
           |||||||||||||||| ||||||||||||||||||||||||||||||  |||| ||||||
Db   40677 TGTGGGCTCCCCACCCATGTTGTACATGCAACCCAGTCCCCTCTGACCACAGAGATTTGT 40736

Qy    2039 qactqaqatgtqacatttqqqattccccat 2068
           |||||||||||||||||||||||||||||
Db   40737 GACTGAGATGTGACATTTGGGATTCCCCAT 40766


Search completed: July 30, 2002, 04:18:32
Job time: 6281 sec
```

```
OM of: US-09-932-678-2 to: GenEmbl:*   out_format : pfs

Date: Jul 30, 2002  6:23 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09932678/runat_29072002_155124_17930/app_query.fasta_1.717
 -DB=GenEmbl -QFMT=fastap -SUFFIX=std.rge -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE 500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09932678_@CGN1_1_4676 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-932-678-2
Query length: 651
Database: GenEmbl:*
Database sequences: 1797656
Database length: 1873333701
Search time (sec): 2289.090000


score_list:
Sequence            Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_pr:AF227156       + 3403.00 5988.00  0.0       2068   ! AF227156 Homo sapiens RRN3 mRN
gb_pr:HSA272050      + 3398.00 5972.63  0.0       3756   ! AJ272050 Homo sapiens mRNA for
gb_pat:AX036047      + 3391.00 5966.93  0.0       2040   ! AX036047 Sequence 1 from Paten
gb_htg:AC012547      + 2039.00 3530.43  2.8e-188  135873  ! AC012547 Homo sapiens clone
gb_pr:AC013444       + 2039.00 3526.60  4.5e-188  192815  ! AC013444 Homo sapiens BAC cl
gb_pat:AX274907      + 1124.00 1962.17  6.2e-101  1461   ! AX274907 Sequence 172 from Pat
gb_pr:BC006441       + 1003.00 1741.80  1.2e-88   2627   ! BC006441 Homo sapiens, Similar
gb_pr:AK055742       + 1000.00 1737.07  2.2e-88   2493   ! AK055742 Homo sapiens cDNA FLJ
gb_pr:BC009198       +  924.00 1610.90  2.3e-81   1177   ! BC009198 Homo sapiens, Similar
gb_in:AY061123       +  783.00 1355.19  4.0e-67   2112   ! AY061123 Drosophila melanogast
gb_pl:SPAC18G6       -  554.50  919.65  7.3e-43   37621  ! Z68198 S.pombe chromosome I c
gb_pl:SCRRN3GEN      +  424.50  718.74  1.1e-31   2672   ! Z71927 S.cerevisiae RRN3 gene.
gb_pl:SCYKL125W      +  424.50  718.06  1.2e-31   2843   ! Z28125 S.cerevisiae chromosome
gb_pr:AC106788       +  398.00  626.39  1.6e-26   170611  ! AC106788 Homo sapiens chromo
gb_htg:AC007615      +  398.00  626.07  1.6e-26   175691  ! AC007615 Homo sapiens chromo
gb_htg:AC009093      +  398.00  624.22  2.1e-26   208008  ! AC009093 Homo sapiens chromo
gb_htg:AC092562      +  393.50  617.40  5.0e-26   187643  ! AC092562 Papio hamadryas clo
gb_htg:AC040158      +  392.00  618.80  4.2e-26   129473  ! AC040158 Homo sapiens chromo
gb_htg:AC092137      -  392.00  617.65  4.9e-26   143900  ! AC092137 Homo sapiens chromo
gb_htg:AC097268      -  390.50  609.07  1.5e-25   247331  ! AC097268 Pan troglodytes clo
gb_htg:AC009130      +  380.50  594.19  9.8e-25   191496  ! AC009130 Homo sapiens chromo
gb_pr:AC092375       -  372.00  580.26  5.9e-24   173166  ! AC092375 Homo sapiens chromo
gb_htg:AC017077      -  372.00  579.89  6.2e-24   179150  ! AC017077 Homo sapiens chromo
gb_pr:HUAF001549     +  368.00  571.50  1.8e-23   202004  ! AF001549 Human Chromosome 16
gb_htg:AC017809      -  366.50  606.82  1.9e-25   6287   ! AC017809 Drosophila melanogast
gb_in:AE003783       +  366.50  576.99  8.9e-24   96018  ! AE003783 Drosophila melanogas
gb_in:AC018484       +  366.50  570.56  2.0e-23   172748  ! AC018484 Drosophila melanoga
gb_in:AC011757       +  366.50  570.33  2.1e-23   176424  ! AC011757 Drosophila melanoga
gb_htg:AC092137      +  341.50  528.36  4.6e-21   143900  ! AC092137 Homo sapiens chromo
gb_htg:AC017077      -  332.00  509.16  5.4e-20   179150  ! AC017077 Homo sapiens chromo
gb_htg:AC022642      -  326.50  508.49  5.8e-20   78347   ! AC022642 Homo sapiens clone H
gb_htg:AC009130      -  280.00  416.49  7.8e-15   191496  ! AC009130 Homo sapiens chromo
gb_htg:AC107769      +  234.50  349.93  4.0e-11   53814   ! AC107769 Mus musculus clone R
gb_htg:AC107769      -  234.00  349.04  4.4e-11   53814   ! AC107769 Mus musculus clone R
gb_pl:AC003096       -  231.50  341.30  1.2e-10   72919   ! AC003096 Arabidopsis thaliana
gb_htg:AC023752      +  229.00  327.33  7.2e-10   174531  ! AC023752 Drosophila melanoga
gb_in:CEC36E8        -  197.50  290.26  8.4e-08   31808   ! Z35597 Caenorhabditis elegans
gb_pl:CNS019XT       +  197.00  330.83  4.6e-10   720   ! AL112233 Botrytis cinerea strai
gb_htg:AC097269      -  182.00  242.93  3.6e-05   196413  ! AC097269 Pan troglodytes clo
gb_htg:AC017733      -  169.50  250.02  1.5e-05   13630   ! AC017733 Drosophila melanogas
gb_pl:T518           -  162.00  215.38  0.0012    96183   ! AC007060 Arabidopsis thali
gb_htg:AC068150      -  152.00  196.43  0.0141    108040  ! AC068150 Homo sapiens chr
gb_htg:AC092119      -  152.00  191.34  0.0270    171930  ! AC092119 Homo sapiens chr
gb_pr:AC092375       -  152.00  191.26  0.0273    173166  ! AC092375 Homo sapiens chr
gb_htg:AC092285      -  152.00  184.67  0.0636    316296  ! AC092285 Homo sapiens chr

seq_name: gb_pr:AF227156

seq_documentation_block:
LOCUS       AF227156             2068 bp    mRNA    linear   PRI 29-APR-2000
DEFINITION  Homo sapiens RRN3 mRNA, complete cds.
ACCESSION   AF227156
VERSION     AF227156.1  GI:7670099
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2068)
  AUTHORS   Moorefield,B., Greene,E.A. and Reeder,R.H.
  TITLE     RNA polymerase I transcription factor Rrn3 is functionally
            conserved between yeast and human
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (9), 4724-4729 (2000)
  MEDLINE   20243763
REFERENCE   2  (bases 1 to 2068)
  AUTHORS   Moorefield,B., Greene,E.A. and Reeder,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2000) Basic Sciences, Fred Hutchinson Cancer
            Research Center, 1100 Fairview Avenue N., Seattle, WA 98109, USA
FEATURES             Location/Qualifiers
     source          1..2068
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16p12"
     CDS             69..2024
                     /note="similar to yeast RNA polymerase I transcription
                     factor Rrn3"
                     /codon_start=1
                     /product="RRN3"
                     /protein_id="AAF66160.1"
                     /db_xref="GI:7670100"
                     /translation="MAAPLLHTRLPGDAAASSSAVKKLGASRTGISNMRALENDFFNS
                     PPRKTVRFGGTVTEVLLKYKKGETNDFELLKNQLLDPDIKDDQIINWLLEFRSSIMYL
                     TKDFEQLISIILRLPWLNRSQTVVEEYLAFLGNLVSAQTVFLRPCLSMIASHFVPPRV
                     IIKEGDVDVSDSDDEDDNLPANFDTCHRALQIIARYVPSTPWFLMPILVEKFPFVRKS
                     ERTLECYVHNLLRISVYFPTLRHEILELIIEKLLKLDVNASRQGIEDAEEIATQTCGG
                     TDSTEGLFNMDEDEETEHETKAGPERLDQMVHPVAERLDILMSLVLSYMKDVCYVDGK
                     VDNGKTKDLYRDLINIFDKLLLPTHASCHVQFFMFYLCSFKLGFAEAFLEHLWKKLQD
                     PSNPAIIRQAAGNYIGSFLARAKFIPLITVKSCLDLLVNWLHIYLNNQDSGTKAFCDV
                     ALHGPFYSACQAVFYTFVFRHKQLLSGNLKEGLQYLQSLNFERIVMSQLNPLKICLPS
                     VVNFFAAITNKYQLVFCYTIIERNNRQMLPVIRSTAGGDSVQICTNPLDTFFPFDPCV
                     LKRSKKFIDPIYQVWEDMSAEELQEFKKPMKKDIVEDEDDDFLKGEVPQNDTVIGITP
                     SSFDTHFRSPSSSVGSPPVLYMQPSPL"
BASE COUNT      573 a    437 c    485 g    573 t
ORIGIN


alignment_scores:
          Quality:  3403.00          Length:    651
            Ratio:     5.227            Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-932-678-2 x AF227156   ..

 Align seq 1/1  to: AF227156  from: 1  to: 2068

        1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSe 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||:
       69 ATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGATGCGGCCGCTTC 118

       17 rSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnM 34
```


119 GTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATA 168
44 eLArqALaLeuGluAsnAspPheAsnSerProProArqLysThrVal 50
169 TGCGTGCATTAGAGAATGACTTTTCAATTCTCCCCCAAGAAAAACTGTT 218
51 ArqPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGlyGl 67
219 CGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGA 268
67 uThrAsnAspPheGluLeuLeuLysAsnGluLeuLeuAspProAspIleL 84
269 AACAAATGACTTTGACTTGTTGAAGAACGAGCTGTTAGATCCAGACATAA 318
84 ysAspAspGlnIleIleAsnTrpLeuLeuGluPheArqSerSerIleMet 100
319 ACGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATG 368
101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArqLeuPr 117
369 TACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCC 418
117 eTrpLeuAsnArqSerGlnThrValValGluGluTyrLeuAlaPheLeuG 134
419 TTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTG 468
134 lyAsnLeuValSerAlaGlnThrValPheLeuArqProCysLeuSerMet 150
469 GTAATCTTCTATCAGCACAGACTGTTTCCTCACACCGTGTCTCAGCATG 518
151 IleAlaSerHisPheValProProArqValIleIleLysGlnGlyAspVa 167
519 ATTGCTTCCCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCGATGT 568
167 IAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPheA 184
569 AGATGTTCAGATTCTGATGATGACGAAGATGATAATCTTCCTGCAAATTTTG 618
184 spThrCysHisArqAlaLeuGlnIleIleAlaArqTyrValProSerThr 200
619 ACACATGTCACAGAGCCTTGCAAATAATAGCAAGATATGTACCATCGACA 668
201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArqLy 217
669 CCGTGGTTTCTCATGCCAATACTGGTGGAAAAATTTCCATTTGTTCGAAA 718
217 sSerGluArqThrLeuGlnCysTyrValHisAsnLeuLeuArqIleSerV 234
719 ATCAGAGACAACACTGGAAGTTACGTTCATAACTTACTAAGGATTAGTG 768
234 alTyrPheProThrLeuArqHisGluIleLeuGluLeuIleIleGluLys 250
769 TATATTTTCCAACCTTGAGGCATCAAATTCTGGAGCTTATTATTGAAAAA 818
251 LeuLeuLysLeuAspValAsnAlaSerArqGlnGlyIleGluAspAlaGl 267
819 CTACTCAAGTTGGATGTGAATGCACGCCAGGCAGGGTATTGAAGATGCTGA 868
267 uGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuP 284
869 AGAAACAGCAACTCAAACTTGTGGTGGGACAGATTCCACGGAAGGATTGT 918
284 heAsnMetAspGluAspGluGluPheGluHisGlnThrLysAlaGlyPro 300
919 TTAATATGGATGAAGATGAAGAAACTGAACATGAAACAAGGCTGGTCCT 968
301 GluArqLeuAspGlnMetValHisProValAlaGluArqLeuAspIleLe 317
969 GAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCGCCTGGACATCCT 1018
317 uMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyL 334
1019 GATGICTTIGGTTTTGTCCTACATGAAGGATGTCTGCTATGTAGATGGTA 1068
334 ysValAspAsnGlyLysThrLysAspLeuTyrArqAspLeuIleAsnIle 350
1069 AGGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATC 1118
351 PheAspLysLeuLeuProThrHisAlaSerCysHisValGlnPhePh 367
1119 TTTGACAAACTCCTGTTGCCCACCCATGCCTCCTGCCATGTACAGTTTTT 1168
367 eMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuG 384
1169 CATGTTTTACCTCTGTAGTTTCAAATTGGGATTCGCAGAGGCATTTTTGG 1218
384 luHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArq 400
1219 AACATCTCTGGAAAAAATTGCAGGACCCAAGTAATCCTGCCATCATCAGG 1268
401 GluAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArqAlaLysPheI 417
1269 CAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAGAGCTAAATTTAT 1318
417 eProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuH 434
1319 TCCTCTTATTACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGC 1368
435 isIleTyrLeuAsnAsnGluAspSerGlyThrLysAlaPheCysAspVal 450
1369 ACATATACCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTT 1418
451 AlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPh 467
1419 GCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACTT 1468
467 eValPheArqHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuG 484
1469 TGTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGTTTGC 1518
484 lnTyrLeuGlnSerLeuAsnPheGluArqIleValMetSerGlnLeuAsn 500
1519 ATTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAAT 1568
501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleTh 517
1569 CCCTGAAGATTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGCAATCAC 1618
517 rAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArqAsnAsnA 534
1619 AAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAATC 1668
534 rqGlnMetLeuProValIleArqSerThrAlaGlyGlyAspSerValGln 550
1669 GCCAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAG 1718
551 IleCysThrAsnProLeuAspThrPhePheProPheAspProCysValIe 567
1719 ATCTGCACAAACCCGCTGGACACCTTCTTCCCCTTTGATCCCTGTGTGCT 1768
567 uLysArqSerLysLysPheIleAspProIleTyrGlnValTrpGluAspM 584
1769 GAAGAGGTCAAAGAAATTCAATTGATCCTATTTATCAGGTATGGGAAGACA 1818
584 etSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle 600
1819 TCAGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAAGGACATA 1868
601 ValGluAspGluAspAspAspPheLeuLysGlyGluValProGlnAsnAs 617
1869 GTGGAAGATGAAGATGATGACTTCTGAAAGGCGAAGTGCCCCAGAATGA 1918
617 prhrValIleGlyIleThrProSerSerPheAspThrHisPheArqSerP 634
1919 TACCGTGATTCGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAGTC 1968

```
     634 roSerSerSerValGlySerProProValLeuTyrMetGlnProSerPro 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1969 CTTCAAGTAGTGTGGGCTCCCCACCCGTGTTGTACATGCAACCCAGTCCC 2018

     651 Leu 651
         |||
    2019 CTC 2021

seq_name: qb_pr:HSA272050

seq_documentation_block:
LOCUS       HSA272050              3756 bp    mRNA    linear   PRI 07-SEP-2000
DEFINITION  Homo sapiens mRNA for transcription initiation factor IA protein
            (TIF-IA gene).
ACCESSION   AJ272050
VERSION     AJ272050.1  GI:10046713
KEYWORDS    TIF-IA; transcription initiation factor.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3756)
  AUTHORS   Bodem,J., Hoffmann-Rohrer,U., Ross,W., Delius,H., Vingron,M. and
            Grummt,I.
  TITLE     Cloning and functional characterization of transcription initiation
            factor TIF-IA, a growth-dependent regulator of ribosomal RNA
            synthesis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3756)
  AUTHORS   Bodem,J.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2000) Bodem J.J., Molekulare Biologie der Zelle
            II, Deutsches Krebsforschungszentrum, INF.280, 69120 Heidelberg,
            GERMANY
FEATURES             Location/Qualifiers
     source          1. .3756
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            23. .3756
                     /gene="TIF-IA"
     CDS             23. .1978
                     /gene="TIF-IA"
                     /function="transcription initiation factor"
                     /codon_start=1
                     /evidence=experimental
                     /product="transcription initiation factor IA protein"
                     /protein_id="CAC07955.1"
                     /db_xref="GI:10046714"
                     /translation="MAAPLLHTRLPGDAAASSSAVKKLCASRTGISNMRALENDFFNS
                     PPRKTVRFGGTVTEVLLKYKKGETNDFELLKNQLLDPDIKDDQIINWLLEFRSSIMYL
                     TKDFEQLISTILRLPWLNRSQTVVEEYLAFLGNLVSAQTVFLRPCLSMIASHFVPPRV
                     IIKEGDVPVSDSDDEDDNLPANFDTCHRALQIIARYVPSTPWFLMPILVEKFPFVRKS
                     ERTLECYVHNLLRISVYFPTLRHEILELIIEKLLKLDVNASKQGIEDAEETATQTCGG
                     TDSTEGLFNMDEDEETEHETKAGPERLDQMVHPVAERLDILMSLVLSYMKDVCYVDGK
                     VDNGKIKDLYRDLINIFDKLLLPTHASCHVQFFMFYLCSFKLGFAEAFLEHLWKKLQD
                     PSNPAIIRQAAGNYIGSFLARAKFIPLITVKSCLDLLVNWLHIYLNNQDSGTKAFCDV
                     ALHGPFYSACQAVFYXFVFRHKQLLSGNLKEGLQYLQSLNFERIVMSQLNPLKICLPS
                     VVNFFAAITNKYQLVFCYTIIERNNRQMLPVIRSTAGGDSVQICTNPLDIFPPFDPCV
                     LKRSKKFIDPIYQVWEDMSAEELQEFKKPMKKDIVEDEIDDFLKGEVPQNDTVIGITP
                     SSFDTHFRSPSSSVGSPPVLYMQPSPL"
     3'UTR           1979. .3756
                     /gene="TIF-IA"
                     /evidence-experimental
BASE COUNT     1060 a     692 c     844 g    1156 t       4 others
ORIGIN


alignment_scores:
            Quality: 3398.00              Length:    651
              Ratio:  5.220                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 99.846
```

```
alignment_block:
 US-09-932-678-2 x HSA272050     .

 Align seq 1/1  to: HSA272050  from: 1  to: 3756

       1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      23 ATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGATGCGGCCGCTTC 72

      17 rSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnM 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      73 GTCCTCTGCAGTTAAGAAGCTGGGCGCTCGAGGACTGGGATTTCAAATA 122

      34 etArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrVal 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     123 TGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTT 172

      51 ArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGlyGl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     173 CGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGA 222

      67 uThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     223 AACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAA 272

      84 ysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     273 AGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATG 322

     101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuPr 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     323 TACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCC 372

     117 oTrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLeuG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     373 TTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTG 422

     134 lyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMet 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     423 GTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATG 472

     151 IleAlaSerHisPheValProProArgValIleIleLysGluGlyAspVa 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     473 ATTGCTTCCCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCGATGT 522

     167 lAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPheA 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      52 AGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCTGCAAATTTTG 572

     184 spThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     573 ACACATGTCACAGAGCCTTGCAAATAATAGCAAGATATGTACCATCGACA 622

     201 ProTrpPheLeuMetProIleLeuValGluLysProPheValArgLy 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     623 CCGTGGTTTCTCATGCCAATACTGGTGCAAAAATTTCCATTTGTTCGAAA 672

     217 sSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerV 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     673 ACAGAGACAACACTGGAATGTTACGTTCATAACTTACTAAGCATTAGTG 722

     234 alTyrPheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLys 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     723 TATATTTTCCAACCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAAA 772

     251 LeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGl 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     773 CTACTCAAGTTGGATGTGAATGCATCCCGGCAGGGTATTGAAGATGCTGA 822

     267 uGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuP 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     823 AGAAACAGCAACTCAAACTTGTGGTGGGACAGATTCCACGGAAGGATTGT 872
```

```
  284 heAsnMetAspGluAspGluGluThrGluHisGlnThrLysAlaGlyPro  300
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  871 TTAATATGGATGAAGATGAAGAAACTGAACATGAAACAAAGGCTGGTCCT  922

  301 GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLe  317
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  923 GAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCGCCTGGACATCCT  972

  317 uMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyL  334
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  973 GATGTCTTGGTTTTGTCCTACATGAAGGATGTCTGCTATGTAGATGGTA  1022

  334 ysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIle  350
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1023 AGGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATC  1072

  351 PheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePh  367
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1073 TTTGACAAACTTCTTCTTCCCACCCATGCCTCCTGCCATGTACAGTTTTT  1122

  367 eMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuG  384
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1123 CATGTTTTACTTCTGTAGTTTCAAATTGGGATTCGCAGAGGCATTTTTGG  1172

  384 luHisLeuTrpLysLysLeuGluAspProSerAsnProAlaIleIleArg  400
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1173 AACATCTTCTGGAAAAAATTGCAGGACCCAAGTAATCCTGCCATCATCAGG  1222

  401 GluAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIl  417
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1223 CAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAAGAGCTAAATTTAT  1272

  417 eProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuH  434
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1273 TCCTCTTATTACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGC  1322

  434 isIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspVal  450
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1323 ACATATACCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTT  1372

  451 AlaLeuHisGlyProPheTyrSerAlaCysGluAlaValPheTyrThrPh  467
      ||||||||||||||||||||||||||||||||||||||||||||||:::||
 1373 GCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACRCCTT  1422

  467 eValPheArgHisLysGluLeuLeuSerGlyAsnLeuLysGluGlyLeuG  484
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1423 TGTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGTTTGC  1472

  484 InTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGluLeuAsn  500
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1473 AGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAAT  1522

  501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleTh  517
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1523 CCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGCAATCAC  1572

  517 rAsnLysTyrGluLeuValPheCysTyrThrIleIleGluArgAsnAsnA  534
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1573 AAATAAGTACTACGTCTGTTCTTCTGCTACACCATCATTGAGAGGAACAATC  1622

  534 rgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGln  550
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1623 GCCAGATGCTGCCAGTCATTAGGAGTACGGCTGGAGGAGACTCAGTGCAG  1672

  551 IleCysThrAsnProLeuAspThrPhePheProPheAspProCysValIe  567
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1673 ATCTGCACAAACCCGCTGGACACCTTCTTCCCCTTTGATCCCTGTGTGCT  1722

  567 eLysArgSerLysLysPheIleAspProIleTyrGlnValTrpGluAspM  584
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1723 GAAGAGGTCAAAGAAATTCATTGATCCTATTTATCAGGTGTGGGAAGACA  1772

  584 etSerAlaGluGluLeuGluGluPheLysLysProMetLysLysAspIle  600
```

```
      ||:||||||||||||||||||||||||||||||||||||||||||||||||
 1773 TGAGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAAGGACATA  1822

  601 ValGluAspGluAspAspAspPheLeuLysGlyGluValProGlnAsnAs  617
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1823 GTGGAAGATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCAGAATGA  1872

  617 pThrValIleGlyIleThrProSerSerPheAspThrHisPheArgSerP  634
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1873 TACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAGTC  1922

  634 roSerSerSerValGlySerProProValLeuTyrMetGluProSerPro  650
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1923 CTTCAAGTAGTGTGGGCTCCCCACCCGTGTTGTACATGCAACCCAGTCCC  1972

  651 [os: 651]
      |||
 1973 CTC 1975

seq name: gb_pat:AX036047

seq_documentation_block:
LOCUS       AX036047                2040 bp    DNA     linear   PAT 15-NOV-2000
DEFINITION  Sequence 1 from Patent WO0055416.
ACCESSION   AX036047
VERSION     AX036047.1  GI:11191586
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2040)
  AUTHORS   Grummt,I. and Vingron,M.
  TITLE     Rna polymerase i transcription factor tif-ia
  JOURNAL   Patent: WO 0055416-A 1 21-SEP-2000;
            DEUTSCHES KREBSFORSCH (DE) ; GRUMMT INGRID (DE) ; VINGRON MARTIN
            (DE)
FEATURES             Location/Qualifiers
     source          1..2040
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             40..1995
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC16268.1"
                     /db_xref="GI:11191587"
                     /translation="MAAPLLHTRLPGDAAASSSAVKKLGASKTGISNMRALENDFFNS
                     PPRKTVREGGTVTEVLLKYKKGETNDFELLKNQLLDPDIKDDQIINWLLEFRSSIMYL
                     TKDEEQLISTLLRLPWLNRSQTVVEEYLAFLGNLVSAQFVFLRPCLSMIASHFVPPRV
                     IIKEGDVDVSDSDDEDDNLPANFDTCHRALQIIARYVPSTPWFLMPILVEKFPFVRKS
                     ERTLECYVHXLLRISVYFPTLRHEILELIIEKLLKLDVNASRQGIEDAEETATQTCGG
                     TDSTEGLFNMDEDEETEHETKAGPERLDQMVHPVAERLDILMSLVLSYMKDVCYVDGK
                     VDNGKTKDLYRDLINIFDKLLLPTHASCHVQFFMFYLCSFKLGFAEAFLEHLWKKLQD
                     PSNPAIIRQAAGNYIGSFLARAKFIPLITVKSCLDLLVNWLHIYLNNQDSGTKAFCDV
                     ALHGPFYSACQAVFYXFVFRHKQLLSGNLKEGLQYLQSLNFERIVMSQLNPLKICLPS
                     VVNFFAAITNKYQLVFCYTIIERNNRQMLPVIRSTAGGDSVQICTNPLDTFFPFDPCV
                     LKRSKKFIDPIYQVWEDMSAEELQEFKKPMKKDIVEDEDDDFLKGEVPQNDIVIGTTP
                     SSFDTHFRSPSSSVGSPPVLYMQPSPL"
BASE COUNT       563 a    434 c    474 g    568 t      2 others
ORIGIN


alignment scores:
            Quality:  3391.00             Length:    651
              Ratio:  5.217                 Gaps:      0
 Percent Similarity:  99.846   Percent Identity: 99.693

alignment block:
 US-09-932-678-2 x AX036047   ..

 Align seq 1/1  to: AX036047  from: 1  to: 2040

       1 M-tAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAlaSe 17
```

```
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 40 ATGGCGGCACCGCTGCTTCACACGCGTTTGCCGGGAGATGCGGCCGCTTC 89

 17 rSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnM 34
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 90 GTCCTCTGCAGTTAAGAAGCTGGGCGCGTCGAGGACTGGGATTTCAAATA 139

 34 etArgAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrVal 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
140 TGCGTGCATTAGAGAATGACTTTTTCAATTCTCCCCCAAGAAAAACTGTT 189

 51 ArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGlyGl 67
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
190 CGGTTTGGTGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGA 239

 67 uThrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleL 84
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
240 AACAAATGACTTTGAGTTGTTGAAGAACCAGCTGTTAGATCCAGACATAA 289

 84 ysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMet 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
290 AGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCTATCATG 339

101 TyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArgLeuPr 117
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
340 TACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTAAGATTGCC 389

117 oTrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPheLeuG 134
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
390 TTGGTTGAATAGAAGTCAAACAGTAGTGGAAGAGTATTTGGCTTTTCTTG 439

134 lyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMet 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
440 GTAATCTTGTATCAGCACAGACTGTTTTCCTCAGACCGTGTCTCAGCATG 489

151 IleAlaSerHisPheValProProArgValIleIleLysGluGlyAspVa 167
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
490 ATTGCTTCCCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCGATGT 539

167 lAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPheA 184
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
540 AGATGTTTCAGATTCTGATGATGAAGATGATAATCTTCCTGCAAATTTTG 589

184 spThrCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
590 ACACATGTCACAGAGCCTTGCAAATAATAGCAAGATATGTACCATCGACA 639

201 ProTrpPheLeuMetProIleLeuValGluLysPheProPheValArgLy 217
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
640 CCGTGGTTTCTCATGCCAATACTCGTGGAAAAATTTCCATTTGTTCGAAA 689

217 sSerGluArgThrLeuGluCysTyrValHisAsnLeuLeuArgIleSerV 234
    ||||||||||||||||||||||||||||||||  ||||||||||||||||
690 ATCAGAGAGAACACTGGAATGTTACGTTCATNACTTACTAAGGATTAGTG 739

234 alTyrPheProThrLeuArgHisGluIleLeuGluLeuIleIleGluLys 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
740 TATATTTTCCAACCTTGAGGCATGAAATTCTGGAGCTTATTATTGAAAAA 789

251 LeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGl 267
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
790 CTACTCAAGTTGGATGTGAATGCATCCCGGCAGGGTATTGAAGATGCTGA 839

267 uGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGluGlyLeuP 284
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
840 AGAAACAGCAACTCAAACTTGTGGTGGGACAGATTCCACGGAAGGATTGT 889

284 heAsnMetAspGluAspGluGluThrGluHisGluThrLysAlaGlyPro 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
890 TTAATATGGATGAAGATGAAGAAACTGAACATGAAACAAAGGCTGGTCCT 939

301 GluArgLeuAspGlnMetValHisProValAlaGluArgLeuAspIleLe 317
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
940 GAACGGCTCGACCAGATGGTGCATCCTGTAGCCGAGCGCCTGGACATCCT 989

317 uMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspGlyL 334
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
990 GATGTCTTTGGTTTTGTCCTACATGAAGGATGTCTGCTATGTAGATGGTA 1039

334 ysValAspAsnGlyLysThrLysAspLeuTyrArgAspLeuIleAsnIle 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1040 AGGTTGATAACGGCAAAACAAAGGATCTATATCGCGACCTGATAAACATC 1089

351 PheAspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePh 367
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1090 TTTGACAAACTGCTGTTGCCCACCCATGCCTCCTGCCATGTACAGTTTTT 1139

367 eMetPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuG 384
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1140 CATGTTTTACCTCTGTAGTTTCAAATTGGGATTCGCAGAGGCATTTTTGG 1189

384 luHisLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArg 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1190 AACATCTCTGGAAAAAATTGCAGGACCCAAGTAATCCTGCCATCATCAGG 1239

401 GlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIl 417
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1240 CAGGCTGCTGGAAATTATATTGGAAGCTTTTTGGCAACAGCTAAATTTAT 1289

417 eProLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuH 434
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1290 TCCTCTTATTACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGC 1339

434 isIleTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspVal 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1340 ACATATACCTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTT 1389

451 AlaLeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPh 467
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1390 GCTCTCCATGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACNCCTT 1439

467 eValPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuG 484
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1440 TGTTTTTAGACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGTTTGC 1489

484 lnTyrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsn 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1490 AGTATCTTCAGAGTCTGAATTTTGAGCGGATAGTGATGAGCCAGCTAAAT 1539

501 ProLeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleTh 517
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1540 CCCCTGAAGATTTGCCTGCCCTCAGTGGTTAACTTTTTTGCTGCAATCAC 1589

517 rAsnLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnA 534
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1590 AAATAAGTACCAGCTCGTCTTCTGCTACACCATCATTGAGAGGAACAATC 1639

534 rgGlnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGln 550
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1640 GCCAGATGCTGCCAGTCATTAGGAGTACCGCTGGAGGAGACTCAGTGCAG 1689

551 IleCysThrAsnProLeuAspThrPhePheProPheAspProCysValLe 567
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1690 ATCTGCACAAACCCGCTGGACACCTTCTTCCCCTTTGATCCCTGTGTGCT 1739

567 uLysArgSerLysLysPheIleAspProIleTyrGlnValTrpGluAspM 584
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1740 GAAGAGGTCAAAGAAATTCATTGATCCTATTTATCAGGTGTGGGAAGACA 1789

584 etSerAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIle 600
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1790 TGAGTGCTGAAGAGCTACAGGAGTTCAAGAAACCCATGAAAAAGGACATA 1839

601 ValGluAspGluAspAspAspPheLeuLysGlyGluValProGlnAsnAs 617
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1840 GTGGAACATGAAGATGATGACTTTCTGAAAGGCGAAGTGCCCCAGAATGA 1889
```

```
     617 pThrValIleGlyIleThrProSerSerPheAspThrHisPheArgSerP 634
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1890 TACCGTGATTGGGATCACACCAAGCTCCTTTGACACGCATTTCCGAAGTC 1939

     634 roSerSerSerValGlySerProProValLeuTyrMetGlnProSerPro 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1940 CTTCAAGTAGTGGGGCTCCCCACCCGTGTTGTACATGCAACCCAGTCCC 1989

     651 Leu 651
         |||
    1990 CTC 1992

Seq name: gb htg:AC012547

Seq documentation block:
LOCUS       AC012547               135873 bp    DNA     linear   HTG 30 MAR 2000
DEFINITION  Homo sapiens clone RP11 11K1, WORKING DRAFT SEQUENCE, 14 unordered
            pieces.
ACCESSION   AC012547
VERSION     AC012547.2  GI:7341888
KEYWORDS    HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 135873)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-11K1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 135873)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boqushavkiy,L., Boukhqalter,B.,
            Brown,A., Castle,A., Colanqelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzBuqh,W., Forrest,C., Funke,R., Gaqe,D.,
            Galagan,J., Gardyna,S., Grant,G., Haqos,B., Hoaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGirk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29 OCT 1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 30, 2000 this sequence version replaced gi:6139119.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
               Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www.seq.wi.mit.edu
            Contact: sequence submissions@genome.wi.mit.edu
                  Project Information
            Center project name: L3178
            Center clone name: 11 K 1
                  Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 127596 bases at least Q40
            Consensus quality: 140929 bases at least Q30
            Consensus quality: 142388 bases at least Q20
            Insert size: 134000; agarose-1p
            Insert size: 134573; sum of contigs
            Quality coverage: 4.9 in Q20 bases; agarose-1p
            Quality coverage: 4.9 in Q20 bases; sum-of-contigs
               . .
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1516: contig of 1516 bp in length
            *     1517  1616: gap of       100 bp
            *     1617     8136: contig of 1520 bp in length
            *     8137  8236: gap of       100 bp
            *     8237     9148: contig of 1912 bp in length
            *     9149  9248: gap of       100 bp
            *     9249     7742: contig of 2494 bp in length
            *     7743  7842: gap of       100 bp
            *     7843    10650: contig of 2808 bp in length
            *    10651 10750: gap of       100 bp
            *    10751    16194: contig of 5444 bp in length
            *    16195 16294: gap of       100 bp
            *    16295    25751: contig of 9457 bp in length
            *    25752 25851: gap of       100 bp
            *    25852    37209: contig of 11358 bp in length
            *    37210 37309: gap of       100 bp
            *    37310    47828: contig of 10519 bp in length
            *    47829 47928: gap of       100 bp
            *    47929    64239: contig of 16311 bp in length
            *    64240 64339: gap of       100 bp
            *    64340    77846: contig of 13507 bp in length
            *    77847 77946: gap of       100 bp
            *    77947    95311: contig of 17365 bp in length
            *    95312 95411: gap of       100 bp
            *    95412   108930: contig of 13519 bp in length
            *   108931 109030: gap of       100 bp
            *   109031   135873: contig of 26843 bp in length.
FEATURES             Location/Qualifiers
     source          1. .135873
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone="RP11 11K1"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1516
                     /note "assembly_fragment"
     misc_feature    1617. .8136
                     /note="assembly_fragment"
     misc_feature    8237. .9148
                     /note "assembly_fragment"
     misc_feature    9249. .7742
                     /note "assembly_fragment"
     misc_feature    7843. .10650
                     /note="assembly_fragment"
     misc_feature    10751. .16194
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    16295. .25751
                     /note "assembly_fragment"
     misc_feature    25852. .37209
                     /note="assembly_fragment"
     misc_feature    37310. .47828
                     /note "assembly_fragment"
     misc_feature    47929. .64239
                     /note "assembly_fragment"
     misc_feature    64340. .77846
                     /note="assembly_fragment"
     misc_feature    77947. .95311
                     /note "assembly_fragment"
     misc_feature    95412. .108930
                     /note "assembly_fragment"
     misc_feature    109031. .135873
                     /note "assembly_fragment"
BASE COUNT    42279 a  25994 c  25146 g  41147 t    1307 others
ORIGIN
```

```
alignment_scores:
           Quality: 2039.00              Length:    645
             Ratio:  3.776                Gaps:     17
 Percent Similarity:  83.721   Percent Identity: 69.612

alignment_block:
 US-09-932-678-2 x AC012547   ..

 Align seq 1/1  to: AC012547  from: 1  to: 135873

      12 GlyAspAlaAlaAlaSerSerSerAlaValLysLysLeuGlyAlaSerAr 28
         |||:::|||   |||||||:::|||:::|||||||||||||||||   |||
   96772 GGGGAAGCAGAGGCTTCAGCTTCTGTGGTCAAGAAGCTGGGTATGTCATG 96821

      28 qThrGlyIleSerAsnMetArgAlaLeuGluAsnAspPhePheAsnSerP 45
          |||:::|||||||:::|||   |||:::::::|||||||||   :::|||||
   96822 GACCAGTATTTCAGATATGCTTGCATTCAAGAATGATTTTTCCAGCTCTC 96871

      45 roProArgLysThrValArgPheGlyGlyThrValThrGluValLeuLeu 61
         ||||||||||||   |||:::::::   ||||||   ||||||||||||
   96872 CCCCAAGAAAAATTGTTCATCTTGATCGAACGGGGACAGAAGTCTTGCCG 96921

      62 LysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLe 78
         |||   |||:::   :::       :::|||||||||||||||:::||||||
   96922 AAGGACAAAGAGTGAAACAAATTTAACTTTGAGCTGTTGAAGAGCCAGTT 96971

      78 uLeuAspProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluP 95
         |   |||   :::|||||||||||||||||||:::|||||||||   :::|
   96972 GTGAGATTCAAACATAAAGGATGACCAGATCTTGAACTGGCTGTGAAAAT 97021

      95 heArgSerSerIleMetTyrLeuThrLysAspPheGluGlnLeuIleSer 111
         ||   ||||||||||||||||||||:::|||||||||||||||||||:::|||
   97022 TTTGCTCTTCTATCATGTACTTGGCAAAAGACTTTGAGCAGCTCCTCAGT 97071

     112 IleIleLeuArgLeuProTrpLeuAsnArgSerGlnThrValValGluGl 128
         |||   ||||||||||||||||||||||||   |||||||||||||||||
   97072 ATTACATTGAGATTGCCTTGGTTGAATAGAATTCAAACAGTGGTGGAAGA 97121

     128 uTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuA 145
         ||||||||||   |||||||:::|||||:::::      :::||||||
   97122 GTATTTGGCTTCTCTGGGTAGTCTTGTAACAACAGA....CTCTTCCTTA 97167

     145 rqProCysLeuSerMetIleAlaSerHisPheVal...ProProArgVal 160
           |||   ||||||:::|||||||||||||:::::   ||||||   :::
   97168 GCCCAAGTCTCAGCATAATTGCTTCCCATTTGGCGTACCCTCCTTCAATG 97217

     161 IleIleLysGluGlyAspValAspValSerAspSerAspAspGluAspAs 177
         ||||||||||::::::   |||:::|||||||||||||||||||   |||||
   97218 ATCATTAAGAAGAGCCATGTAAATGTTTCAGATTCTGATGATGGAGATGA 97267

     177 pAsnLeuProAlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleA 194
         |   ||||||||:::||||||||   |||   :::|||||||||   |
   97268 TTATCTTCCTGCAGATTTTGACACACGTCATACAACCTTACAAATAACAG 97317

     194 laArgTyrValProSerThrProTrpPheLeuMetProIleLeuValGlu 210
         ||||||||||||||   :::   |||||||||:::|||||||||||||||
   97318 CAAGATATGTCCCACTGGCACTGTGGTTTCTTTTGCCCATACTGGTGGAA 97367

     211 LysPheProPheValArgLysSerGluArgThrLeuGluCysTyrValHi 227
            ||||||||||||   :::|||||||||||||||||||||||||:::||
   97368 ACATTTCCATTTGTT...CAATCAGAGAGAACACTGGAATGTTACATTCA 97414

     227 sAsnLeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIleL 244
         ||||||||||:::||||||:::|||||||||||||||||||:::|||||||
   97415 TAACTTACTAAAGATTAGTATATACTTTCCAACCTTGAGGAATGAAATTC 97464

     244 euGluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArg 260
         ||:::|||||||||||:::::|||:::|||||||||||||||||||:::
   97465 TGAAGCTTATTATTAAAAATCTATTCAAATTGGATGTGAATACATTCTGG 97514

     261 GlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyTh 277
         :::   |||||||||||||||||||||||||||:::|||||||||||||||||
```

```
   97515 CATGATATTGAAGATGCTGAGGAAACAGCAGCTCAAACTTGTGGTGGGAC 97564

     277 rAspSerThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluR 294
         |||||||   |||   ||||:|||||||:::|||||||||||||||||
   97565 AGATTCCATGGAATGATTGTTCAATATGAACGAAGATGAAGAGACT.... 97610

     294 isGluThrLysAlaGlyProGluArgLeuAspGlnMetValHisProVal 310
         .   :::|||||||      :::   |||   ||||||||||||||||||
   97611 ..GTTCTAAGGCTGACTCCAAATGGCTCGGTCAGATGGTGCACCCTGTA 97658

     311 AlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAs 327
         ||||||   ||||||   ||:::|||||||:::|||:::|||:::||||
   97659 GCTGAATGCCTGGACACCCTGCTGTCTTTGCTTTTGCAGTACTTGAAGTA 97708

     327 pValCysTyrVal...Asp..GlyLysValAspAsnGlyLysThrLysAs 342
         ||||||||||||         ||||||||||||||||:::|||   ||||||
   97709 TGTCTGCTATGTAAAATACATGGTAAGGTTGATAACAACAAAATAAAGGA 97758

     342 pLeu...Tyr..ArgAspLeuIleAsnIlePheAspLysLeuLeuLeuPr 357
         ||||   |||  :::   ||||||||||||||||||
   97759 TTTATAATATATCACAAACTGATAAACATATTTGACAA............ 97796

     357 oThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSerP 374
         |||   ||||||||||||||||||||||||||||||||||||||||||:::::
   97797 .ACCCCTGCCTCCTGTCACGTACAGTTTTTTATGTTTTACCTCTGTGGCT 97845

     374 heLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrp...LysLys 389
         ::||||||   ::::::||||||||||||||||||:::|||||   ||||||
   97846 TAAAATTGCGACTCTCAGAAGCATTTTTGGAACGTCTCTGGGAAAAAAAA 97895

     390 LeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTy 406
         :::||||||||:::||||||:::|||||||   ||||||||||||||||
   97896 AIGCAGGATCCAAATAATCCTGTCATCATCAGGTGAGCTGCTGGAAATTA 97945

     406 rIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle...ThrV 422
         ||||||||||   ||:|||:::||||||||||||||||||||   |||:
   97946 TATTGGAAGCTCTTTGGCAAAAGCTAAATTCATTCCTCTCATTACAACTA 97995

     422 alLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsn 438
         ::||||||||||||   |||   ||||||||   |||   |||||||:::
   97996 TAAAATCAIG[illegible]AGGTCTTTCGGTTAACTGGCCACACAAATACCTTACT 98045

     439 AsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyPr 455
         |||||||||           ||||||||||||||:::|||||||||
   98046 AACCAGGATCTGAGAATGAAGGCTTTTGTGATATTGCTCTCCATCAAAC 98095

     456 oPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisL 472
         ||||||||||||   ||||||||||||:::|||||||||||||||||||
   98096 ATTTTACTCAGCCTTCCAAGCTGTGTTCCACACCTTTGTTTTTAGACACA 98145

     472 ysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSer 488
         ||||||||||||||   ||||||||||||||||||||:::   ||||||
   98146 AGCAACTTTTAAGCGTAAACCTGAAAGAAGGTTTGAGATGTCTTCAGTGC 98195

     489 LeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCy 505
         |||:::|||      ||||||      ||||||||||||||||:::||||||
   98196 CTACATTTTTAAGGGATAGTGACTAGGCAGCTAAATCCTCTGAATATTTG 98245

     505 s.LeuProSerValValAsnPhePheAlaAlaIlefhrAsnLysTyrGlnL 522
         |||||||||||||||||||:::||||||||||:::|||||||||||||||
   98246 CCTGCCCTCAGTGGTTAACTTATTTGCTGCTGTCACGAATAAATATCAGC 98295

     522 euValPheCysTyrThrIleIleGluArgAsnAsn.ArgGlnMetLeuPr 538
         |||||   ||||||||||||||||||||||||||
   98296 TAGTC...TGCTACACTATCATTGAGAGGAACAACGGCC........... 98331

     538 oValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnP 555
                 |||:::|||||||||||||||||||||      ||||||||||
   98332 .......AGAAATACAGCTGGAGGAGACTCGGTGTGAACGTGCACCAACC 98374

     555 roLeuAspThrPhePheProPheAspProCysValLeuLysArgSerLys 571
         |||||||:::|||||||||||||||||||||:::|||||||:::|||:::
   98375 CATTGGAC[illegible]CTTCTCCCCTTTGATCCTTGTGCGCTTAAGCAGTCAAGG 98424
```

```
     572 LysPheIleAspProIleTyrGlnValTrpGluAspMetSerAlaGlnG 588
         ||| |||||| |||||||||||:::||||||||||:::|||:::||||
   98425 AAAGTCATTGATTCTATTTATCAGATATGGCAAGACATAAGTGTCGAAGA 98474

     588 nLeuGlnGluPheLysLysProMetLysLysAspIleValGluAspGluA 605
         ||||||||||||||||||||||||||||||:::  :::|||  :::
   98475 GCTTCAGGAGTTTAAGAAACCTATGAAAAAGGAG...ATGGAAAAAGATG 98521

     605 spAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIleGly 621
         ||||||:::|||||||||||||||||||  |||||||||||:::||||||
   98522 GAGATGATCTTTGAAAGGTCAAGTTCCCCCAAATGATACCATGATCGGG 98571

     622 IleThrProSerSerPheAspThr...His..PheArgSerProSerSer 636
         |||  ||||||  ||||||:::  ||| |||:::|||||||||||||
   98572 ATTATACCGAGCCACTTTGATGCGATGCACATTTCCAAAGTCCTTCAAGT 98621

     637 SerValGlySerProProValLeuTyrMet 646
         |||||||||||||||||||||||||:::::
   98622 AGTCTGGGCTCCCCACCTGTGTTACACCTG 98651

seq name: gb pr:AC014444

seq documentation block:
LOCUS       AC014444               192815 bp    DNA     linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens BAC clone RP11 470B22 from 2, complete sequence.
ACCESSION   AC014444
VERSION     AC014444.8  GI:18094267
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 192815)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 192815)
  AUTHORS   Waliqorski,J., Haakenson,W. and Spalding,L.
  TITLE     The sequence of Homo sapiens BAC clone RP11-470B22
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 192815)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 192815)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jan 9, 2002 this sequence version replaced gi:13518264.
                    Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
                    Summary Statistics
            Center project name: H_NH0470B22
           ---  -  .

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >
            40); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI 11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is AC024616; the clone sequenced to
            the right is RP11-190J24.  Actual start of this clone is at base
            position 1 of RP11-470B22; actual end is at base position 192815 of
            RP11 470B22.

            Unresolved homopolymeric run from base position 164727 to 164754.

FEATURES             Location/Qualifiers
     source          1. .192815
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "2"
                     /map "2"
                     /clone "RP11-470B22"
                     /clone_lib "RPCI 11"
     repeat_region   1. .575
                     /rpt_family "L1"
     repeat_region   611. .878
                     /rpt_family "ERVL"
     repeat_region   879. .1032
                     /rpt_family "(TAGA)n"
     repeat_region   1046. .1092
                     /rpt_family "ERVL"
     repeat_region   1987. .2346
                     /rpt_family "L2"
     repeat_region   2737. .2972
                     /rpt_family "L1"
     repeat_region   3702. .3749
                     /rpt_family "AT_rich"
     misc_feature    4870. .4891
                     /note "match to EST BE071256 (NID:g8416143)"
     repeat_region   6880. .6992
                     /rpt_family "MIR"
     repeat_region   8677. .8934
                     /rpt_family "Alu"
     repeat_region   8773. .8800
                     /rpt_family "AT rich"
     repeat_region   8955. .8989
                     /rpt_family "AT_rich"
     repeat_region   9207. .9247
                     /rpt_family "L2"
     repeat_region   9342. .9773
                     /rpt_family "L2"
     repeat_region   9977. .10011
                     /rpt_family "(TTTA)n"
     repeat_region   10976. .11265
                     /rpt_family "Alu"
     repeat_region   12178. .12207
                     /rpt_family "(T)n"
     repeat_region   12362. .12390
                     /rpt_family "AT_rich"
     repeat_region   12702. .12741
                     /rpt_family "(TG)n"
     repeat_region   12731. .12758
```

```
                     /rpt_family="(GA)n"
     repeat_region   12758. .12785
                     /rpt_family="(TG)n"
     repeat_region   13033. .13054
                     /rpt_family="(CAAAA)n"
     repeat_region   13195. .13699
                     /rpt_family="L1"
     repeat_region   13543. .13567
                     /rpt_family="AT_rich"
     repeat_region   14142. .14305
                     /rpt_family="MIR"
     repeat_region   14500. .14534
                     /rpt_family="AT_rich"
     repeat_region   14617. .14735
                     /rpt_family="MIR"
     repeat_region   14736. .14836
                     /rpt_family="MER2_type"
     repeat_region   14837. .15106
                     /rpt_family="Alu"
     repeat_region   15107. .15317
                     /rpt_family="MER2_type"
     repeat_region   15318. .15385
                     /rpt_family="MIR"
     repeat_region   15578. .16066
                     /rpt_family="ERV1"
     repeat_region   16072. .16112
                     /rpt_family="AT_rich"
     repeat_region   18192. .18222
                     /rpt_family="AT_rich"
     repeat_region   18612. .18899
                     /rpt_family="Alu"
     misc_feature    19181. .19485
                     /note="match to EST R89091 (NID:q953918) yq02f04.s1"
     misc_feature    19879. .21714
                     /note="similar to Rattus norvegicus EST Y16973
                     (NID:q4727035)"
     misc_feature    19890. .20284
                     /note="match to EST R89197 (NID:q954024) yq02f04.r1"
     misc_feature    19915. .20369
                     /note="similar to Homo sapiens EST BG535886
                     (NID:g13527431)"
     misc_feature    19923. .20378
                     /note="similar to Homo sapiens EST AA213789 (NID:g1812416)
                     zr91a11.s1"
     misc_feature    19923. .20369
                     /note="similar to Homo sapiens EST BG428305
                     (NID:g13334811)"
     misc_feature    20091. .20378
                     /note="similar to Homo sapiens EST AI742153 (NID:g5110441)
                     wq49f01.x1"
     misc_feature    20158. .20664
                     /note="similar to Homo sapiens EST AV703279
                     (NID:g10720608)"
     misc_feature    20243. .21062
                     /note="similar to Mus musculus EST BG295610
                     (NID:g13057417)"
     misc_feature    20412. .21044
                     /note="similar to Homo sapiens EST BE734354
                     (NID:g10148346)"
     misc_feature    20567. .20666
                     /note="similar to Homo sapiens EST BG778634
                     (NID:g14048951)"
     misc_feature    20592. .20963
                     /note="similar to Homo sapiens EST AA481295 (NID:g2210847)
                     zv02e05.r1"
     misc_feature    20703. .21362
                     /note="similar to Homo sapiens EST BG502871
                     (NID:g13464388)"
     misc_feature    20798. .21056
                     /note="similar to Homo sapiens EST AV649508
                     (NID:g9870522)"
     misc_feature    20862. .21269
                     /note="similar to Homo sapiens EST AA191111 (NID:g1779805)
                     zp86a03.r1"
     misc_feature    20930. .21593
                     /note="similar to Homo sapiens EST BF203684
                     (NID:g11097270)"
     misc_feature    20930. .21038

alignment_scores:
            Quality: 2049.00              Length:    645
              Ratio:  3.776                Gaps:      17
 Percent Similarity:  83.721   Percent Identity: 69.612

alignment_block:
 US-09-932-678-2 x AC013444   ..

 Align seq 1/1  to: AC013444  from: 1  to: 192815

      12 GlyAspAlaAlaAlaSerSerSerAlaValLysLysLeuGlyAlaSerAr 28
         |||:::|||  |||||||:::|||:::|||||||||||||||  |||
   19844 GGGGAAGCAGAGGCTTCAGCTTCTGTGGTCAAGAAGCTGGGTATGTCAIG 19893

      28 qThrGlyIleSerAsnMetArqAlaLeuGluAsnAspPheAsnSerP 45
         |||:::|||||||:::|||  |||:::::::|||||||  :::|||||
   19894 GACCAGTATTTCAGATAIGCTTGCATTCAAGAATGATTTTTCCAGCTCTC 19943

      45 roProArqLysThrValArqPheGlyGlyThrValThrGluValLeuLeu 61
         ||||||||||:  |||:::::  |||:||  ||||||||||||
   19944 CCCCAAGAAAAATTGTTCATCTTGATGGAACGGGGACAGAAGTCTTGCCC 19993

      62 LysTyrLysLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLe 78
         |:|  |||:::  :::      :::||||||||||||||||||:||||
   19994 AAGGACAAAGAGTGAAACAAATTTAACTTTGAGCTGTTGAAGAGCCAGTT 20043

      78 uLeuAspProAspIleLysAspAspGlnIleIleAsnTrpLeuLeuGluP 95
         |  |||  :::||||||||||||||||:::|||||||||  :::|
   20044 GTGAGATTCAAACATAAAGGATGACCAGATCTTGAACTGGCTGTGAAAAT 20093

      95 heArqSerSerIleMetTyrLeuThrLysAspPheGluGlnLeuIleSer 111
         ||  ||||||||||||||||||:::|||||||||||||||||||:|||
   20094 TTTGCTCTTCTATCATGTACTTGGCAAAAGACTTTGAGCAGCTCCTCAGT 20143

     112 IleIleLeuArqLeuProTrpLeuAsnArqSerGlnThrValValGluGl 128
         |||  ||||||||||||||||||||||||||  |||||||||||||||||
   20144 ATTACATTGAGATTGCCTTGGTTGAATAGAATTCAAACAGTGGTGGAAGA 20193

     128 uTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuA 145
         ||||||||||  ||||||:::||||||:::::      :::|||||||
   20194 GTATTGGCTTCTCTGGGTAGTCTTGTAACAACAGA....CTCTTCCTTA 20239

     145 rqProCysLeuSerMetIleAlaSerHisPheVal...ProProArqVal 160
         |||  ||||||:::|||||||||||:::::::  ||||||  :::
   20240 GCCCAAGTCTCAGCATAATTGCTTCCCATTGGCGTACCCTCCTTGAATG 20289

     161 IleIleLysGluGlyAspValAspValSerAspSerAspAspGluAspAs 177
         ||||||||||:::::  |||:::||||||||||||||||  ||||||
   20290 ATCATTAAGAAGAGTCATGTAAATGTTTCAGATTCTGATGATGGAGATGA 20339

     177 pAsnLeuProAlaAsnPheAspThrCysHisArqAlaLeuGlnIleIleA 194
         |  ||||||||||:::||||||||  |||  :::||||||||||  |
   20340 TTATCTTCCTGCACATTTTGACACACGTCATACAACCTTACAAATAACAG 20389

     194 laArqTyrValProSerThrProTrpPheLeuMetProIleLeuValGlu 210
         ||||||||||||:  :::  ||||||||||:::|||||||||||||||
   20390 CAAGATATCTCCCACTGGCACTGTGGTTTCTTTTGCCCATACTGGTGGAA 20439

     211 LysPheProPheValArqLysSerGluArqThrLeuGluCysTyrValHi 227
         ||||||||||||  :::||||||||||||||||||||||||||:::||
   20440 ACATTTCCATTTGTT...CAATCAGAGAGAACACTGGAATGTTACATTCA 20486

     227 sAsnLeuLeuArqIleSerValTyrPheProThrLeuArqHisGluIleL 244
         |.||||||||:::.|||||:::||||||||||||||||||:::||||||
   20487 TAACTTACTAAAGATTAGTATATACTTTCCAACCTTGAGGAATGAAATTC 20536

     244 euGluLeuIleIleGluLysLeuLysLeuAspValAsnAlaSerArq 260
```

```
              ||:::|||||||||:::::|||:::|||||||||||||||||:::
  20537 TGAAGCTTATTATTAAAAATCTATTCAAATTGGATGTGAATACATTCTGG 20586

    261 GlnGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyTh 277
        :::   ||||||||||||||||||||||:::|||||||||||||||
  20587 CATGATATTGAAGATGCTGAGGAAACAGCAGCTCAAACTTGTGGTGGGAC 20636

    277 rAspSerThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluH 294
        |||||||    |||    ||||||||||||:::|||||||||||||
  20637 AGATTCCATGGAATGATTGTTCAATATGAACGAAGATGAAGAGACT.... 20682

    294 isGluThrLysAlaGlyProGlnArgLeuAspGlnMetValHisProVal 310
             :::||||||       :::    |||    ||||||||||||||||
  20683 ..GTTTCTAAGGCTGACTCCAAATGGCTCGGTCAGATGGTGCACCCTGTA 20730

    311 AlaGlnArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAs 327
        ||||||    ||||||    |||:::||||||:::|||:::|||:::|||
  20731 GCTGAATGCCTGGACACCCTGCTGTCTTTGCTTTTGCAGTACTTGAAGTA 20780

    327 pValCysTyrVal...Asp..GlyLysValAspAsnGlyLysThrLysAs 342
         ||||||||||||           ||||||||||||||:::|||    |||||
  20781 TGTCTGCTATGTAAAATACATGGTAAGGTTGATAACAACAAAATAAAGGA 20830

    342 pLeu...Tyr..ArgAspLeuIleAsnIlePheAspLysLeuLeuLeuPr 357
        ||||    |||   :::    |||||||||||||||||
  20831 TTTATAATATATCACAAACTGATAAACATATTTGACAA............ 20868

    357 oThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCysSerP 374
         |||    ||||||||||||||||||||||||||||||||||||||::::
  20869 .ACCCCTGCCTCCTGTCACGTACAGTTTTTTATGTTTTACCTCTGTGGCT 20917

    374 heLysLeuGlyPheAlaGlnAlaPheLeuGluHisLeuTrp...LysLys 389
        ::||||||     ::::::||||||||||||||||:::||||||    ||||||
  20918 TAAAATGGCGACTCTCAGAAGCATTTTTGGAACGTCTCTGGGAAAAAAAA 20967

    390 LeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTy 406
        :::|||||||||:::||||||:::|||||||||    ||||||||||||||
  20968 ATGCAGGATCCAAATAATCCTGTCATCATCAGGTGAGCTGCTGGAAATTA 21017

    406 rIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIle...ThrV 422
        ||||||||||    ||||||:::|||||||||||||||||||||    |||:
  21018 TATTGGAAGCTCTTTGGCAAAAGCTAAATTCATTCCTCTCATTACAACTA 21067

    422 alLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsn 438
        ::|||||||||||||    |||    |||||||||    |||    ||||||:::
  21068 TAAAATCATGCCTAGGTCTTTCGGTTAACTGGCCACACAAATACCTTACT 21117

    439 AsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyPr 455
        |||||||||            |||||||||||||||:::|||||||||
  21118 AACCAGGATCTGAGAATGAAGGCTTTTTGTGATATTGCTCTCCATCAAAC 21167

    455 oPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisL 472
          ||||||||||||    ||||||||||||:::||||||||||||||||||
  21168 ATTTTACTCAGCCTTCCAAGCTGTGTTCCACACCTTTGTTTTTAGACACA 21217

    472 ysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSer 488
        ||||||||||||    |||||||||||||||||||:::     ||||||
  21218 AGCAACTTTTAAGCGTAAACCTGAAAGAAGGTTTGAGATGTCTTCAGTGC 21267

    489 LeuAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCy 505
        |||:::|||       ||||||        |||||||||||||||:::|||||
  21268 CTACATTTTTAAGGATAGTGACTAGGCAGCTAAATCCTCTGAATATTTG 21317

    505 sLeuProSerValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnL 522
        ||||||||||||||||||||:::|||||||||:::||||||||||||||||
  21318 CCTGCCCTCAGTGGTTAACTTATTTGCTGCTCTCACGAATAAATATCAGC 21367

    522 euValPheCysTyrThrIleIleGluArgAsnAsn.ArgGlnMetLeuPr 538
        |||||    ||||||||||||||||||||||||||||
  21368 TAGTC...TGCTACACTATCATTGAGAGGAACAACGGCC........... 21403

    538 oValIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnP 555
                |||:::||||||||||||||||||||||        ||||||||||
  21404 .......AGAAATACAGCTGGAGGAGACTCGGTGTGAACGTGCACCAACC 21446

    555 roLeuAspThrPhePheProPheAspProCysValLeuLysArgSerLys 571
        ||||||:::|||||||||||||||||||||||:::||||||:::|||:::
  21447 CATTGGACGCCTTCTTCCCCTTTGATCCTTGTGCGCTTAAGCAGTCAAGG 21496

    572 LysPheIleAspProIleTyrGlnValTrpGluAspMetSerAlaGluGl 588
        |||    ||||||    |||||||||:::|||||||||:::|||:::|||||
  21497 AAAGTCATTGATTCTATTTATCAGATATGGGAAGACATAAGTGTCGAAGA 21546

    588 uLeuGlnGluPheLysLysProMetLysLysAspIleValGluAspGluA 605
        ||||||||||||||||||||||||||||||:::    :::|||    :::
  21547 GCTTCAGGAGTTTAAGAAACCTATGAAAAAGGAG...ATGGAAAAAGATG 21593

    605 spAspAspPheLeuLysGlyGluValProGlnAsnAspThrValIleGly 621
          ||||||:::||||||||||||||||||    ||||||||:::||||||
  21594 GAGATGATCTTTTGAAAGGTGAAGTTCCCCAAATGATACCATGATCGGG 21643

    622 IleThrProSerSerPheAspThr...His..PheArgSerProSerSer 636
        |||    ||||||    |||||||:::    |||   |||:::||||||||||||
  21644 ATTATACCGAGCCACTTTGATGCGATGCACATTTCCAAAGTCCTTCAAGT 21693

    637 SerValGlySerProProValLeuTyrMet 646
        ||||||||||||||||||||||||::::::
  21694 ACTGTGGGCTCCCCACCTGTGTTACACCTG 21723
```

seq_name: gb pat:AX274907

```
seq_documentation_block:
LOCUS       AX274907                1461 bp    DNA     linear   PAT 29-OCT-2001
DEFINITION  Sequence 172 from Patent WO0172777.
ACCESSION   AX274907
VERSION     AX274907.1  GI:16547539
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Hillman,J.L., Baughn,M.R., Yue,H., Lal,P., Lu,D.A., Patterson,C.,
            Azimzai,Y., Bandman,O., Tang,Y.T., Mathur,P., Shah,P., Au Young,J.
            and Reddy,R.
  TITLE     Transcription factors
  JOURNAL   Patent: WO 0172777-A 172 04-OCT-2001;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..1461
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 2646274CB1"
BASE COUNT      413 a    335 c    298 g    415 t
ORIGIN


alignment_scores:
          Quality: 1124.00              Length:     486
            Ratio:    3.986               Gaps:       8
 Percent Similarity:  58.025   Percent Identity: 51.646

alignment_block:
 US-09-932-678-2 x AX274907 ..

 Align seq 1/1  to: AX274907  from: 1  to: 1461

    195 ArgTyrValProSerThrProTrpPheLeuMetProIleLeuValGluLy 211
        :::|||           :::||||||||||||||||||||||||||||||
    263 AAATACAAAATTACGGCACCGTGGTTTCTCATGCCAATACTGGTGGAAAA 312

    211 sPheProPheValArgLysSerGluArgThrLeuGluCysTyrValHisA 228
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    313 ATTTCCATTTGTTCGAAAATCAGAGAGAACACTGGAATGTTACGTTCATA 362

    228 snLeuLeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeu 244
```

```
    ||||||||||||||||||||||||||||||||||||||||||||||||||
363 ACTTACTAAGGATTAGTGTATATTTTCCAACCTTGAGGCATGAAATTCTG 412

245 GluLeuIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGl 261
    ||||||||||||||||||||||||||||||||||||||||||||||||||
413 GAGCTTATTATTGAAAAACTACTCAAGCTGGATGTGAATGCATCCCGGCA 462

261 nGlyIleGluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrA 278
    |||||||||||||||||||||||||||||||:::||||||||||||||||
463 GGGTATTGAAGATGCTGAAGAAACAGCAAATCAAACTTGTGGTGGGACAG 512

278 spSerThrGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHis 294
    ||||||||||||||||||||||||||
513 ATTCCACGGAAGGATTGTTTAATATG........................ 538

295 GluThrLysAlaGlyProGluArgLeuAspGlnMetValHisProValAl 311

538 .................................................. 538

311 aGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspV 328

538 .................................................. 538

328 alCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyr 344

538 .................................................. 538

345 ArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAlaSe 361

538 .................................................. 538

361 rCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyP 378
                                                ||||
539 ..............................................GGAT 542

378 heAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspProSer 394
    ||||||||||||||||||||||||||||||||:::|||||||||||||||
543 TCGCAGAGGCATTTTTGGAACATCTTTGGAAAAACTTGCAGGATCCAAGT 592

395 AsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLe 411
    ||||||||||||||||||||||||||||||||||||||||||||||||||
593 AATCCTGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCTTTTT 642

411 uAlaArgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspL 428
    ||||||||||||||||||       |||||||||||    ||||||||||
643 GGCAAGAGCTAAATTTATTTCTCTTATTACTGTAAAACCATGCCTAGATC 692

428 euLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThr 444
    ||||||||||||||||||||||||||||||||||||||||||||||||||
693 TTTTGGTTAACTGGCTGCACATATACCTTAATAACCAGGATTCGGGAACA 742

445 LysAlaPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGl 461
    ||||||||||||||||||||||||||||||||||||||||||||||||||
743 AAGGCATTCTGCGATGTTGCTCTCCATGGACCATTTTACTCAGCCTGCCA 792

461 nAlaValPheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyA 478
    ||||||||||||||||||||||||||||||||||||||||||||||||||
793 AGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAGCTTTTGAGCGGAA 842

478 snLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIle 494
    ||||||||||||||||||||||     |||||||||||||||||||||||
843 ACCTGAAAGAAGGTTTGCAGTATCCTCAGAGTCTGAATTTTGAGCGGATA 892

495 ValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValAs 511
    ||||||||||||||||||||||||||||||||||||||||||||||||||
893 GTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCTGCCCTCAGTGGTTAA 942

511 nPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThrI 528
    ||||||||||||||||||||:::
943 CTTTTTTGCTGCAATCACAAAG............................ 964

528 leIleGluArgAsnAsnArgGlnMetLeuProValIleArgSerThrAla 544
                                                :::|||:::
965 ......................................ATGAAGACTTGT 976

545 GlyGlyAspSerValGlnIleCysThrAsnProLeuAspThrPhePhePr 561
    |:|                         |||   ::::::    |||:::
977 GGATATGGATGGTGGTGATGGTTGCACAACAATATCAATTTATTTTATAC 1026

561 oPheAspProCysValLeuLys............................ 568
      :::|||||:::    |||
1027 CACTGAACCGTGCACTTCAAAATGGTTAAGATGGTAAGTTTTATGTTGTG 1076

569 .........................ArgSerLysLysPheIleAspPro 576
                             |||:::|||:::   :::   |||
1077 TATTTTACCATAATAAAAAAAATTGTAGAGGGAAAAACAGTCTGCCTCCA 1126

577 IleTyrGlnValTrpGlu.......................AspMetSe 585
    :::      :::|||:::                       :::   ||
1127 CTTTTG...ATATGGGACTGCTAACATCTTCCACCTTCCGTCTCCCCTTC 1173

585 rAlaGluGluLeuGlnGluPheLysLysProMetLysLysAspIleValG 602
    ||||    :::|||    :::::::::||||||
1174 TGCCCCACATCTGGGCAAGCTAAGAAAGCCT................... 1204

602 luAspGluAspAspAspPheLeuLysGlyGluValProGlnAsnAspThr 618
                                 :::   ::::::    |||
1205 ...........................GCTGCTCTCTCCTCTGGCACC 1225

619 ValIleGlyIleThrProSer............SerPheAspThrHisPh 631
         :::   |||:::                |||   :::||||||:::
1226 AGCTGGAAATTCATACCCAACAAGCCCTAGCCTTCCCACCAGACCCACAT 1275

631 eArgSerProSerSerSerValGlySerProProValLeuTyrMetGlnP 648
    :   |||||||||   |||            |||              :::|
1276 TTCATCCCCATCCCCATCGCATCCCCATCCC...............CATC 1310

648 roSerPro 650
    ||   |||
1311 CCCATCCC 1318
```

seq_name: gb_pr:BC006441

seq_documentation_block:

```
LOCUS       BC006441                 2627 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, Similar to RNA polymerase I transcription factor
            RRN3, clone MGC:13169 IMAGE:3351791, mRNA, complete cds.
ACCESSION   BC006441
VERSION     BC006441.1  GI:13623642
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2627)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
```

```
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 18 Row: d Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.
FEATURES             Location/Qualifiers
     source          1..2627
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:13169 IMAGE:3351791"
                     /tissue_type="Eye, retinoblastoma"
                     /clone_lib="NIH_MGC_16"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             1698..2156
                     /codon_start=1
                     /product="Similar to RNA polymerase I transcription factor
                     RRN3"
                     /protein_id="AAH06441.1"
                     /db_xref="GI:13623643"
                     /translation="MGFAEAFLEHLWKNLQDPSNPAIIRQAAGNYIGSFLARAKFISL
                     ITVKPCLDLLVNWLHIYLNNQDSGTKAFCDVALHGPFYSACQAVFYTFVFRHKQLLSG
                     NLKEGLQYPQSLNFERIVMSQLNPLKICLPSVVNFFAAITKMKTCGYGWW"
BASE COUNT      837 a    492 c    525 g    773 t
ORIGIN

alignment_scores:
            Quality:   1003.00          Length:     458
              Ratio:     3.903            Gaps:       8
 Percent Similarity:   56.114   Percent Identity:  49.782

alignment_block:
 US-09-932-678-2 x BC006441 ..

 Align seq 1/1 to: BC006441 from: 1 to: 2627

     223 GluCysTyrValHisAsnLeuLeuArgIleSerValTyrPheProThrLe 239
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1509 GAATGTTACGTTCATAACTTACTAAGGATTAGTGTATATTTTCCAACCTT 1558

     239 uArgHisGlnIleLeuGluLeuIleIleGluLysLeuLeuLysLeuAspV 256
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1559 GAGGCATGAAATTCTGGAGCTTATTATTGAAAAACTACTCAAGCTGGATG 1608

     256 alAsnAlaSerArgGlnGlyIleGluAspAlaGluGlnThrAlaThrGln 272
         |||||||||||||||||||||||||||||||||||||||||||:::|||
    1609 TGAATGCATCCCGGCAGGGTATTGAAGATGCTGAAGAAACAGCAAATCAA 1658

     273 ThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGluAs 289
         ||||||||||||||||||||||||||||||||||||||||||
    1659 ACTTGTGGTGGGACAGATTCCACGGAAGGATTGTTTAATATG........ 1700

     289 pGluGluThrGlnHisGluThrLysAlaGlyProGluArgLeuAspGlnM 306

    1700 .................................................. 1700

     306 etValHisProValAlaGluArgLeuAspIleLeuMetSerLeuValLeu 322

    1700 .................................................. 1700

     323 SerTyrMetLysAspValCysTyrValAspGlyLysValAspAsnGlyLy 339

    1700 .................................................. 1700

     339 sThrLysAspLeuTyrArgAspLeuIleAsnIlePheAspLysLeuLeuL 356

    1700 .................................................. 1700

     356 euProThrHisAlaSerCysHisValGlnPhePheMetPheTyrLeuCys 372

    1700 .................................................. 1700

     373 SerPheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLy 389
                      ||||||||||||||||||||||||||||||||||::
    1701 .............GGATTCGCAGAGGCATTTTTGGAACATCTTTGGAAAAA 1738

     389 sLeuGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnT 406
         :|||||||||||||||||||||||||||||||||||||||||||||||||
    1739 CTTGCAGGATCCAAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATT 1788

     406 yrIleGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrVal 422
         ||||||||||||||||||||||||||||||||||    |||||||||||
    1789 ACATTGGAAGCTTTTTGGCAAGAGCTAAATTTATTCCTCTTATTACTGTA 1838

     423 LysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAs 439
         |||     |||||||||||||||||||||||||||||||||||||||||
    1839 AAACCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAA 1888

     439 nGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProP 456
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1889 CCAGGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCATGGACCAT 1938

     456 heTyrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLys 472
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1939 TTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAG 1988

     473 GlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLe 489
         |:||||||||||||||||||||||||||||||||||||    ||||||||
    1989 CAGCTTTTGAGCGGAAACCTGAAAGAAGGTTTGCAGTATCCTCAGAGTCT 2038

     489 uAsnPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCysL 506
         :||||||||||||:||||||||||||||||||||||||||||||||||||
    2039 GAATTTTGAGCGGATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCC 2088

     506 euProSerValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeu 522
         |:|||||||||||||||||||||||||||||||||:::
    2089 TGCCCTCAGTGGTTAACTTTTTTGCTGCAATCACAAAG............ 2126

     523 ValPheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProVa 539

    2126 .................................................. 2126

     539 lIleArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProL 556
                :::|||:::|||                          |||   :
    2127 ....ATGAAGACTTGTGGATATGGATGGTGGTGATGGTTGCACAACAATA 2172

     556 euAspThrPhePheProPheAspProCysValLeuLys............ 568
         :::::    |||:::        :::||||||:::     |||
    2173 TCAATTTATTTTATACCACTGAACCGTGCACTTCAAAATGGTTAAGATGG 2222

     569 .........................................ArgSerLy 571
                                                  |||:::||
    2223 TAAGTTTTATGTTGTGTATTTTACCATAATAAAAAAAATTGTAGAGGGAA 2272

     571 sLysPheIleAspProIleTyrGlnValTrpGlu................ 582
         |:::    :::    |||:::          :::|||:::
    2273 AAACAGTCTGTCTCCACTTTTG...ATATGGGACTGCTAACATCTTCCAC 2319

     583 ........AspMetSerAlaGluGluLeuGlnGluPheLysLysProMet 596
                 :::    ||||||     :::|||    :::::::::||||||
    2320 CCTCCCTCTCCCCCTCTGCCCCACATCTGGGCAAGCTAAGAAAGCCT... 2366

     597 LysLysAspIleValGluAspGluAspAspAspPheLeuLysGlyGluVa 613
                                                          ::
    2367 .............................................GCTGC 2371

     613 lProGlnAsnAspThrValIleGlyIleThrProSer............S 626
         :    ::::::    |||          :::    |||:::          |
    2372 TCTCTCCTCTGGCACCAGCTGGAAATTCATACCCAACAAGCCCTACCCCT 2421

     626 erPheAspThrHisPheArgSerProSerSerSerValGlySerProPro 642
```

```
          ||    :::||||||:::   ||||||||    |||          |||
     2422 CCCACCAGACCCACATTTCATCCCCATCCCCATCGCATCCCCATCCC... 2468

      643 ValLeuTyrMetGlnProSerPro 650
                      :::|||   |||
     2469 ...........CATCCCCATCCC 2480

seq_name: gb_pr:AK055742

seq_documentation_block:
LOCUS       AK055742                2493 bp    mRNA    linear   PRI 31-OCT-2001
DEFINITION  Homo sapiens cDNA FLJ31180 fis, clone KIDNE2000266.
ACCESSION   AK055742
VERSION     AK055742.1  GI:16550546
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens kidney cDNA to mRNA, clone_lib:KIDNE2
            clone:KIDNE2000266.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,
            Ono,Y., Hotuta,T., Hiraoka,S., Murakawa,K., Takiguchi,S.,
            Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,
            Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,
            Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y.,
            Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y.,
            Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2493)
  AUTHORS   Isogai,T., Otsuki,T. and Sugiyama,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: RAB and
            HRI.
FEATURES             Location/Qualifiers
     source          1..2493
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="KIDNE2000266"
                     /tissue_type="kidney"
                     /clone_lib="KIDNE2"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      554 a    700 c    761 g    478 t
ORIGIN


alignment_scores:
          Quality: 1000.00              Length:    196
            Ratio:   5.208                Gaps:      0
 Percent Similarity:  97.959   Percent Identity: 95.918

alignment_block:
 US-09-932-678-2 x AK055742   ..

 Align seq 1/1  to: AK055742  from: 1  to: 2493

      374 PheLysLeuGlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLe 390
          ::::::   ||||||||||||||||||||||||||||||||||||||||||||
      703 TATCGCACGGGATTCGCAGAGGCATTTTTGGAACATCTTTGGAAAAAATT 752

      390 uGlnAspProSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrI 407
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      753 GCAGGATCCAAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATTATA 802

      407 leGlySerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLys 423
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      803 TTGGAAGCTTTTTGGCAAGAGCTAAATTTATTCCTCTTATTACTGTAAAA 852

      424 SerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGl 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      853 CCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATACCTTAATAACCA 902

      440 nAspSerGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheT 457
          |||||||||||||||||||||||||::::||||||||||||||||||||||
      903 GGATTCGGGAACAAAGGCATTCTGCAATGTTGCTCTCCATGGACCATTTT 952

      457 yrSerAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGln 473
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      953 ACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTAGACACAAGCAG 1002

      474 LeuLeuSerGlyAsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAs 490
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1003 CTTTTGAGCGGAAACCTGAAAGAAGGTTTGCAGTATCTTCAGAGTCTGAA 1052

      49  nPheGluArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuP 507
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     105  TTTTGAGCGGATAGTGATGAGCCAGCTAAATCCCCTGAAGATTTGCCTGC 1102

      507 roSerValValAsnPhePheAlaAlaIleThrAsnLysTyrGlnLeuVal 523
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1103 CCTCAGTGGTTAACTTTTTTGCTGCAATCACAAATAAGTACCAGCTCGTC 1152

      524 PheCysTyrThrIleIleGluArgAsnAsnArgGlnMetLeuProValIl 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1153 TTCTGCTACACCATCATCGAGAGGAACAATCGCCAGATGCTGCCAGTCAT 1202

      540 eArgSerThrAlaGlyGlyAspSerValGlnIleCysThrAsnProLeuA 557
          ||||||||||:::||||||||||||||||||    ||||||||||||||
     1203 TAGGAGTACCACTGGAGGAGACTCAGTGCAGACCTGCACAAACCCACTGG 1252

      557 spThrPhePheProPheAspProCysValLeuLysArg 569
          ||||||||   ||||||||||||||||||||||||||
     1253 ACACCTTCGTCCCCTTTGATCCCTGTGTGCTGAAGAGG 1290

seq_name: gb_pr:BC009198

seq_documentation_block:
LOCUS       BC009198                1177 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, Similar to RNA polymerase I transcription factor
            RRN3, clone MGC:15321 IMAGE:3678732, mRNA, complete cds.
ACCESSION   BC009198
VERSION     BC009198.1  GI:14327947
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1177)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
```

```
          Greene, Mark Ketteman and Anuradha Madan

          Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
          Series: IRAL Plate: 24 Row: k Column: 14
          This clone was selected for full length sequencing because it
          passed the following selection criteria: matched mRNA gi: 3355302.
FEATURES             location/Qualifiers
     source          1..1177
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "MGC:15421 IMAGE:3678732"
                     /tissue_type "Lymph, Burkitt lymphoma"
                     /clone_lib "NIH_MGC_8"
                     /lab_host "DH10B-R"
                     /note "Vector: pOTB7"
     CDS             316..1647
                     /codon_start=1
                     /product "Similar to RNA polymerase I transcription factor
                     RRN3"
                     /protein_id "AAH09198.1"
                     /db_xref "GI:14327948"
                     /translation "MRALENDFFNSPPRKTVRFGGTVTEVLLKYKKGETNDFELLKNQ
                     LLDPDIKDDQLINWLLEFRSSVMYLTKDFEQLLTSTLRLPWLNRSQTVVEEYLAFLGN
                     LVSA"
BASE COUNT       339 a    222 c    253 g    363 t
ORIGIN


alignment scores:
            Quality:  924.00              Length:    264
              Ratio:  4.714                 Gaps:      2
 Percent Similarity:  74.242   Percent Identity:  73.485

alignment block:
 US-09-932-678-2 x BC009198   ..

 Align seq 1/1  to: BC009198  from: 1  to: 1177

        1 MetAlaAlaProLeuLeuHisThrArg.LeuProGlyAspAlaAlaAlaS 17
          ||||||||||||||||||||||||||   |||  ||||||||||||||||
       46 ATGGCGGCACCGCTGCTTCACACGTTGTTTGTCGGGAGATCGCGGCCGCTT 85

       17 erSerSerAlaValLysLysLeuGlyAlaSerArgThrGly..... 30
          |||||||||||||||||  ||||||||||||||||||||||
       86 CGTCCTCTCGCAGTCAAGACGCTGGGCGCGTCGAGGACTGGCTTCTACTTG 135

       30 ..................................................  30

      136 AATGAAGACAAAAATCTAGGCCGGGCGCGGTTGCTCACGCCTGTAATCCC 185

       30 ..................................................  30

      186 AGCACTTTGGGAGGCCGAGGTGGGCGGATCACCTGAAGTCAGGAGTTCGA 235

       30 ..................................................  30

      236 GACCAGCCTGGCCAACATGGCGAAACCCCGTCTCTACTAAAAATACAAAAA 285

       31 .....................IleSerAsnMetArgAlaLeuGluAsnAs 40
                               ||||||||||||||||||||||||||||||
      286 TTAGCCAGGCGTGGTGGCAGGCATTTCAAATATGCGTGCATTAGAGAATGA 335

       40 pPhePheAsnSerProProArgLysThrValArgPheGlyGlyThrValT 57
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      336 TTTTTCAATTCTCCCCCAAGAAAACGTTCGGTTTGGTGGAACTGTGA 385

       57 hrGluValLeuLeuLysTyrLysLysGlyGluThrAsnAspPheGluLeu 73
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      386 CAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTG 435

       74 LeuLysAsnGlnLeuLeuAspProAspIleLysAspAspGlnIleIleAs 90
          ||||||||||||||||||||||||||||||||||||||||||||||||||

      436 TTGAAGAACCAGCTGTTAGATCCAGACATAAAGGATGACCAGATCATCAA 485

       90 nTrpLeuLeuGluPheArgSerSerIleMetTyrLeuThrLysAspPheG 107
          ||||||||||||||||||||||||||:::||||||||||||||||||||
      486 CTGGCTGCTAGAATTCCGTTCTTCTGTCAGTACTTGACAAAAGACTTTG 535

      107 luGlnLeuIleSerIleIleLeuArgLeuProTrpLeuAsnArgSerGln 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      536 AGCAACTTATCAGTATTATATTGAGATTGCCTTGGTTGAATAGAAGTCAA 585

      124 ThrValValGluGluTyrLeuAlaPheLeuGlyAsnLeuValSerAlaGl 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      586 ACAGTAGTGGAAGAGTATTTGGCTTTCTTGGTAATCTTGTATCAGCATA 635

      140 nThrValPheLeuArgProCysLeuSerMetIleAlaSerHisPheValP 157
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      636 GACTGTTTCCTCAGACCGTGTCTCAGCATGATTGCTTCCCATTTGTGC 685

      157 roProArgValIleIleLysGluGlyAspValAspValSerAspSerAsp 173
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      686 CTCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGATTCTGAT 735

      174 AspGluAspAspAsnLeuProAlaAsnPheAspThrCysHisArgAlaLe 190
          ||||||||||||||||||||||||||||||||||||   ||||||||||
      736 GATGAAGATGATAATCTTCCTGCAAATTTTGACACATATCACAGAGCCTT 785

      190 uCysIleIleAlaArgTyrValProSerThrProTrpPhe 203
          ||||||||||||||||||||||||||||      :::|||
      786 CCAAATAATAGCAAGATATGTACCATCATCAGTATACTTT 825

seq name:: gb_in:AY061123

seq documentation_block:
LOCUS       AY061123                2112 bp    mRNA    linear   INV 08-NOV-2001
DEFINITION  Drosophila melanogaster LD10456 full length cDNA.
ACCESSION   AY061123
VERSION     AY061123.1  GI:16768904
KEYWORDS    FLI_CDNA.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 2112)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
            Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,
            Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,
            Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanenavong,S., Wan,K.,
            Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-2001) Berkeley Drosophila Genome Project,
            Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, USA
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory
            Berkeley, CA 94720
            This clone was sequenced as part of a high throughput process to
            sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
            Science 2000). The sequence has been subjected to integrity checks
            for sequence accuracy, presence of a polyA tail and contiguity
            within 100 kb in the genome. Thus we believe the sequence to
            reflect accurately this particular cDNA clone.  However, there are
            artifacts associated with the generation of cDNA clones that may
            have not been detected in our initial analyses such as internal
            priming, priming from contaminating genomic DNA, retained introns
            due to reverse transcription of unspliced precursor RNAs, and
            reverse transcriptase errors that result in single base changes.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our Web site
            (http://fruitfly.berkeley.edu) or send email to
            cdna@fruitfly.berkeley.edu.
FEATURES             location/Qualifiers
```

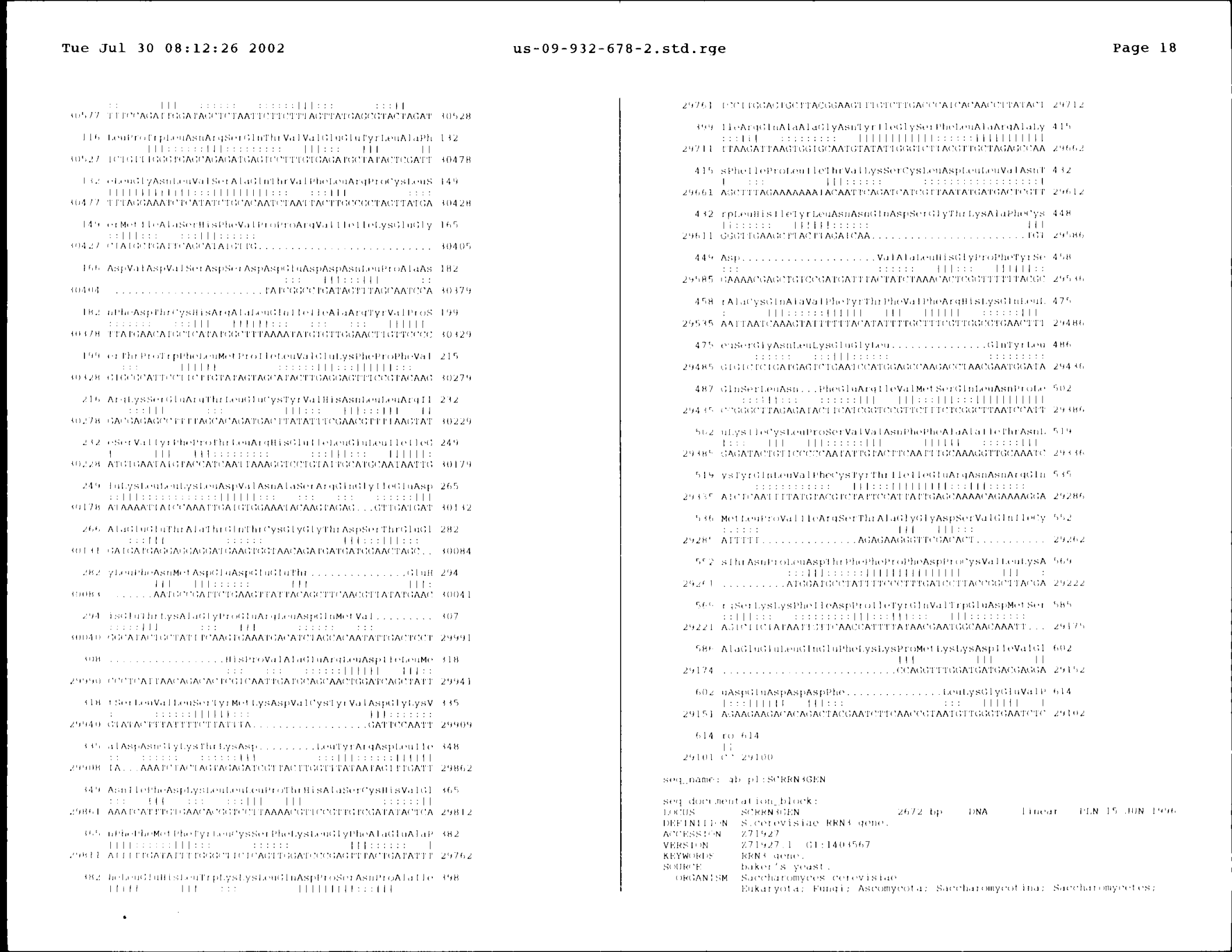
```
     source          1..2112
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /map="40D2-40D3"
     gene            1..2112
                     /gene="CG3278"
                     /note="alignment with genomic scaffold AE003783. gene does
                     not completely overlap longest ORF"
                     /db_xref="FLYBASE:FBgn0032988"
     CDS             178..2013
                     /gene="CG3278"
                     /note="Longest ORF"
                     /codon_start=1
                     /db_xref="FLYBASE:FBgn0032988"
                     /product="LD10456p"
                     /protein_id="AAL28671.1"
                     /db_xref="GI:16768904"
                     /translation="MSFYTSKTDVTSILKTFSGVKRNQANAVAINKVRFSTPKEKGLA
                     ESVRVALEERNFHLVKEFTYFLREAELSDDEVVIIIKDAKRIVHNLTPEFVTVVEALL
                     SLNWKKRSSEIIEAYTEFCIELMVTHNRYLPIGVSKLIVHWIPGDMDESDWVHGCPSE
                     HVRNELKPIHKVLDRILTAVPMAFDIIDAISAKFPYFKKPYHVTAGYLYNVLWLLEY
                     KPIFEELVFQLVLQKLLLLDVNAPKDEIELEMNDEEEKVEADTLFEIDDVCAENVPKS
                     EKSVSHPVGKTLDVCLFLLYKFFDGKCRLNENSSEEQRRIANRLFKMLLYLFDEVLLP
                     SHNTHHVQFILFYATSLRSVYSEAFLDSLWTKVQNPNVSAVIRHAAVGYMASFLSRAR
                     FLPLSTVTFYLKELTRWAHTYIDDSDAYKQNCSLKANLVFYSVCQAVFYLIAFRAKDL
                     TASSKDLLFLQSLQLSRLAMCHLNPLRYCLAPVATAFAGVTRTYQLAYCHTVLERNAR
                     RKLATVYGHEKCMPDETLESFFPFDPYVLKLSKKYIETNYMVYQSNDTDEYVYGSNKY
                     GHSRKRGDSEMLEEDEFLIVDKRPKNFDLSKSQEFDKQFHFGSSP"
BASE COUNT      649 a    405 c    414 g    644 t
ORIGIN

alignment_scores:
           Quality:   783.00              Length:     659
             Ratio:   1.967                 Gaps:      18
 Percent Similarity:  60.395   Percent Identity:  30.501

alignment_block:
 US-09-932-678-2 x AY061123   ..

 Align seq 1/1  to: AY061123  from: 1  to: 2112

     19 SerAlaValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetAr 35
        |||:::|||||:::      |||:::   :::::|||:::::::::||
    229 TCTGGCGTAAAACGAAATCAGGCAAATGCCGTGGCCATCAACAAAGTTCG 278

     35 gAlaLeuGluAsnAspPhePheAsnSerProProArgLysThrValArgP 52
        |                 |||::::::|||   :::|||
    279 A.................TTTTCCACACCCAAAGAGAAG.......... 300

     52 heGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGlyGluThr 68
            |||   :::   :::|||    :::                 |||
    301 ....GGGCTCGCAGAATCAGTGCGCGTTGCCCTA.........GAAGAA 336

     69 AsnAspPheGluLeuLeuLysAsn.........GlnLeuLeuAspProAs 82
        ::::::|||:::|||:::|||:::              |||   :::   ::
    337 CGAAATTTCCATCTGGTCAAGGAGTTCACATACTTCCTCCGGGAAGCAGA 386

     82 pIleLysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerI 99
        ::::::|||||||:::::::::     :::   :::   :::         :
    387 ACTCAGCGATGATGAGGTTGTAATAATCATAAAAGATGCTAAACGAATTG 436

     99 leMetTyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArg 115
        ::       ||||||   :::|||        :::::::::   :::|||
    437 TTCATAATTTGACTCCAGAATTTGTAACCGTTGTCGAGGCGTTACTGTCC 486

    116 LeuProTrpLeuAsnArgSerGlnThrValValGluGluTyrLeuAlaPh 132
        |||    |||    :::|||||||:::    :::::::|||     |||       ||
    487 CTTAACTGGAAGAAAAGAAGTTCAGAGATAATCGAAGCTTATACTGAATT 536

    132 eLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuS 149
        |         :::|||:::    :::::::::    :::|||          :::|

    537 TTGTATCGAACTAATGGTGACCCATAATAGATATTTACCGATTGGAGTCT 586

    149 erMetIleAlaSerHisPheValProProArgValIleIleLysGluGly 165
        ||    :::      |||::::::|||                      |||
    587 CTAAACTAATTGTCCACTGGATTCCT.....................GGA 615

    166 AspValAspValSerAsp......SerAspAspGluAspAspAsnLeuPr 180
        |||:::|||   ||||||                     :::::::::::
    616 GATATGGATGAATCCGATTGGGTCCACGGATGCCCTTCGGAGCACGTGCG 665

    180 oAlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyrV 197
          ::::::                |||::::::|||:::   |||       |
    666 AAATGAGTTGAAACCCATCCATAAAGTACTGGATCGCATCTTAACTGCCG 715

    197 alProSerThrProTrpPheLeuMetProIleLeuValGluLysPhePro 213
        |||||      :::          ::::::::        :::      ||||||||
    716 TGCCAATGGCTTTCGATATTATCATAGATGCCATATCCGCCAAATTCCCC 765

    214 PheValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeuLe 230
        :::      :::|||        ::::::             |||:::::::|||:::||
    766 TATTTCAAAAAGCCGTACCATGTGACGGCAGGGTATCTGTATAACGTGCT 815

    230 uArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLeuI 247
        |       :::          |||    |||     :::::::::    :::::::::|||:
    816 TTGGTTATTAGAATACAAGCCTATTTTCGAAGAACTTGTTTTTCAACTAG 865

    247 leIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIle 263
        ::::::::||||||||    ||||||||||||||    ::::::   |||
    866 TTTTACAAAAGCTTTTATTATTAGATGTTAATGCTCCAAAAGATGAAATT 915

    264 GluAspAlaGluGluThrAlaThrGlnThrCysGlyGlyThrAspSerTh 280
        |||
    916 GAA............................................... 918

    280 rGluGlyLeuPheAsnMetAspGluAspGluGluThrGluHisGluThr. 296
              ||             |||||:::|||:::::::|||    :::|||
    919 .......TTAGAAATGAATGATGAGGAGGAAAAGGTGGAAGCGGACACCT 961

    297 .........................LysAlaGlyProGluArgLeuAsp 304
                                 :::     |||:::
    962 TGTTTGAAATCGACGATGTATGCGCCGAAAATGTACCCAAATCTGAAAAA 1011

    305 GlnMetValHisProValAlaGluArgLeuAspIleLeuMetSerLeuVa 321
        ::::::   ||||||||||||::::::    |||||||:::    :::    |||::
   1012 TCTGTGAGTCATCCTGTTGGCAAAACCTTGGATGTATGCCTATTCTTGCT 1061

    321 lLeuSerTyrMetLysAspValCysTyrVal......AspGlyLysValA 336
        :    :::::::::              |||    :::          ::::::::    :
   1062 TTATAAGTTTTTTGATGGCAAATGTCGCCTTAATGAAAATAGTAGCGAAG 1111

    336 soAsnGlyLysThrLysAsp...LeuTyrArgAspLeuIleAsnIlePhe 351
        :::::   :::|||    :::     |||::::::    |||:::    :::|||
   1112 AACAGCGAAGGACTGCCAACCGATTGTTTAAGATGTTACTTTATTTGTTC 1161

    352 AspLysLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMe 368
        |||::::::||||||||||||:::|||     :::     ||||||||||||:::::
   1162 GACGAGGTGTTATTGCCCAGCCACAACACGCATCATGTGCAATTCATCCT 1211

    368 tPheTyrLeuCysSerPheLysLeuGlyPheAlaGluAlaPheLeuGluH 385
        ::|||||      |||:::::::        :::::::||||||||||||:::
   1212 ATTTATGCAACTAGTCTGCGGTCGGTATATTCCGAAGCTTTTCTTGATT 1261

    385 isLeuTrpLysLysLeuGlnAspProSerAsnProAlaIleIleArgGln 401
           ||||||    |||:::|||:::|||:::          |||:::||||||:::
   1262 CTTATGGACGAAAGTCCAAAATCCGAATGTATCCGCTGTAATTCGACAT 1311

    402 AlaAlaGlyAsnTyrIleGlySerPheLeuAlaArgAlaLysPheIlePr 418
        ||||||    :::|||:::::|||||||||||::::::|||||||:::|||
   1312 GTTGCGGTGGGATACATGGCTAGCTTTTTATCTCGCGCACGATTTTTGCC 1361

    418 oLeuIleThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisI 435
        ||||    ||||||             |||        |||:::::::|||    |||
   1362 GTTAGCACTGTTACTTTCTACCTAAAAGAGTTGACAAGGTGGGCTCATA 1411
```

```
 435 leTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAla 451
       |||:::::::::::::|||:::        |||        |||:::::::
1412 CATACATAGATGACTCCGATGCA...TATAAGCAGAACTGCTCTCTGAAA 1458

 452 LeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheVa 468
        :::        |||||||||:::||||||||||||||||    :::::
1459 GCCAATTTGGTTTTCTATTCTGTTTGCCAAGCTGTGTTTTATTTAATTGC 1508

 468 lPheArgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGlyLeuGlnT 485
     :||||||     |||:::|||     :::::::::        :::   |||     :
1509 ATTTCGCGCAAAAGATTTGACGGCCAGTTCA......AAAGATTTGCTCT 1552

 485 yrLeuGlnSerLeuAsnPheGluArgIleValMetSerGlnLeuAsnPro 501
     ::||||||||||||:::::::::::|||::::::::|||     :::|||||||||
1553 TCCTCCAATCCCTCCAACTATCGCGCCTTGCTATGTGTCATTTAAATCCC 1602

 502 LeuLysIleCysLeuProSerValValAsnPhePheAlaAlaIleThrAs 518
     |||:::     ||||||        |||::::::   ||||||::::::|||::
1603 CTGCGGTATTGTTTAGCCCCAGTGGCCACAGCATTTGCTGGTGTTACAAG 1652

 518 nLysTyrGlnLeuValPheCysTyrThrIleIleGluArgAsnAsnArgG 535
     :     |||||||||:::::::|||:::|||::::::|||||||||     |||:
1653 AACTTACCAGCTGGCCTATTGTCATACTGTTTGGAGCGCAACGCTCGCC 1702

 535 lnMetLeuProValIleArgSerThrAlaGlyGlyAspSerValGlnIle 551
     ::     |||                 :::|||:::     |||     :::          :::
1703 GTAAACTG............GCTACAGTTTACGGTCATGAAAAATGCATG 1740

 552 CysThrAsnProLeuAspThrPhePheProPheAspProCysValLeuLy 568
           :::     |||::::::||||||||||||||||||     ||||||||
1741 CCCGATGAGACGCTGGAATCGTTTTTTCCCTTCGATCCTTATGTACTTAA 1790

 568 sArgSerLysLysPheIleAspProIleTyrGlnValTrpGluAspMetS 585
     |     |||||||||:::|||:::         |||:::|||:::::::::     :
1791 GTTATCAAAAAAATATATTGAGACTAATTATATGGTTTACCAGAGCAATG 1840

 585 erAlaGluGluLeu............GlnGluPheLysLysProMetLys 597
     :::::::::|||                  :::::::::                |||
1841 ACACGGATGAGTATGTCTATGGGTCTAACAAATACGGCCATTCTAGGAAA 1890

 598 Lys...AspIleValGluAspGluAspAspAspPheLeuLysGlyGluVa 613
     :::     |||                  |||:::|||:::||||||
1891 CGTGGTGACTCTGAAATGCTAGAGGAAGACGAATTCCTA........... 1929

 613 lProGlnAsnAspThrValIleGlyIleThrProSerSerPheAspThr. 629
                       ::::::               |||::::::||||||
1930 ................ATTGTCGACAAACGTCCAAAGAATTTTGATCTGT 1963

 630 ...........................HisPheArgSerProSerSer 636
                                |||||| 
1964 CTAAAAGTCAGGAATTTGACAAACAATTTCATTTC.............. 1998

 637 SerValGlySerProProValLeuTyr 645
           ||||||     |||     :::|||
1999 ......GGATCTTCGCCATGATTTTAC 2019
```

seq_name: gb_pl:SPAC18G6

```
seq_documentation_block:
LOCUS       SPAC18G6               37621 bp    DNA     linear   PLN 04-APR 2000
DEFINITION  S.pombe chromosome I cosmid c18G6.
ACCESSION   Z68198
VERSION     Z68198.1  GI:3702195
KEYWORDS    40S ribosomal protein; chromosome segregation; GCN1 homologue;
            heterochromatin protein; Homol D box; mal3; mating factor; mral;
            protein transport; rab subfamily; rna polymerase i specific
            transcription initiation factor; rps7; serine
            hydroxymethyltransferase; shm2; translational activator; tRNA-Val;
            vesicular traffic; YPT1 related protein 3; ypt3.
SOURCE      fission yeast.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (bases 1 to 37621)
  AUTHORS   Connor,R., Churcher,C.M., Barrell,B.G., Rajandream,M.A., Walsh,S.V.
            and Wood,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC 1995) Schizosaccharomyces pombe chromosome 1
            sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
            CB10 1RQ E-mail: barrell@sanger.ac.uk
COMMENT     On Oct 6, 1998 this sequence version replaced gi:1122365.
            Notes:
            Details of yeast sequencing at the Sanger Centre are available on
            the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_pombe)
            Protein coding regions (CDS) have been predicted with the help of
            computer analysis using the Genefinder program in PomBase (an ACEDB
            database) with additional predictions for the branch acceptor sites
            supplied by the program Sp3splice. CAUTION: It is possible that for
            any individual CDS we may have underestimated or overestimated the
            number of introns/exons or we may not have chosen the correct
            splice donor/acceptor sites.
            CDS are numbered using the following system eg SPAC5H10.01c. SP (S.
            pombe), A (chromosome I), c5H10 (cosmid name), .01 (first CDS), c
            (complementary strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
            Cosmid c18G6 overlapped by c1327 at the 5' end and c13G6 at the 3'
            end.
FEATURES             Location/Qualifiers
     source          1..37621
                     /organism="Schizosaccharomyces pombe"
                     /strain="972h-"
                     /db_xref="taxon:4896"
                     /chromosome="I"
                     /map="1L"
                     /clone="cosmid c18G6"
     gene            complement(1..193)
                     /gene="SPAC1327.01c"
                     /note="SPAC18G6.16c; SPAC1783.09c"
     misc_feature    1..98
                     /note="nominal overlap with cosmid SPAC1327, EM:AL035254
                     S. pombe chromosome 1"
     CDS             complement(1..193)
                     /partial
                     /gene="SPAC1783.09c"
                     /note="SPAC18G6.16c, len:64 aa"
                     /codon_start=1
                     /label=SPAC18G6.16c
                     /product="putative trancription factor"
                     /protein_id="CAB59875.1"
                     /db_xref="GI:6179673"
                     /translation="MMQQRSGSLSLLSNAVQAGQSSDPMQDDPAKTEPGTPKGYDGSK
                     NSSPASVPSWIHEKTKAGKD"
     gene            complement(1994..2773)
                     /gene="SPAC18G6.01c"
     CDS             complement(1994..2773)
                     /gene="SPAC18G6.01c"
                     /note="SPAC18G6.01c, len:259aa; similarity: to
                     hypothetical protein SW:YHZ8_YEAST P38884  (27.1% identity
                     in 192 aa overlap)"
                     /codon_start=1
                     /label=SPAC18G6.01c
                     /product="hypothetical protein"
                     /protein_id="CAA92381.1"
                     /db_xref="GI:1122366"
                     /db_xref="SWISS-PROT:Q10102"
                     /translation="MFGTRFFAYQFCKNWLSKRPGLVFWSSVALINAMYQTRPLYCES
                     VTKTIQQTYEGLVQKRINNEDELFLLGSGPRYVSFLSIHVYDIELYIQKKDVQTVHII
```

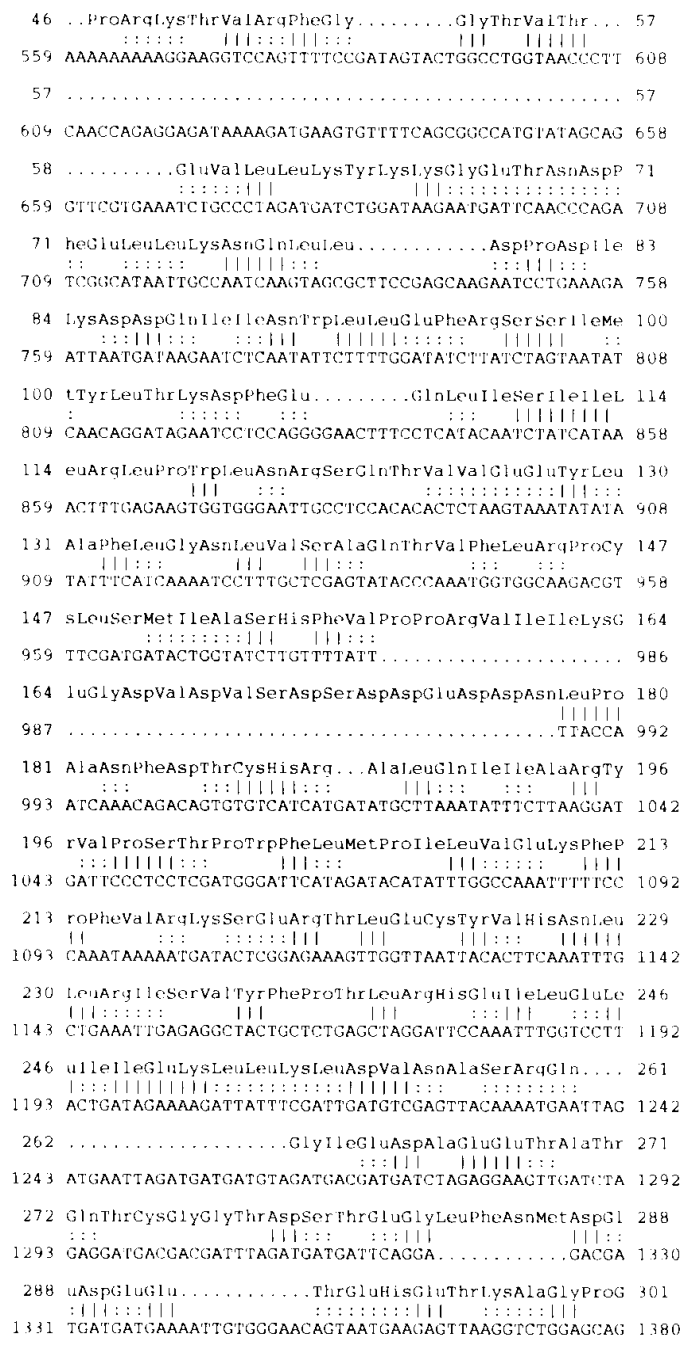
```
                LQKEVDPKLGLEMSMKDEEIGSRIVAALLKNEMKYAIKIVPTRNTNFSHLRGGFVRGL
                QSRMSQNDPAEQAAVSKFRSTFPVNRTCFKETALWMKLYGNDFCYIYKDSEIGHMHDE
                NYMVSNLFLKGYLVGPRVNSEQARESVCLTLRRIMDGTLLF"
gene            complement(3075. .5957)
                /gene="SPAC18G6.02c"
                /note="chp1"
CDS             complement(3075. .5957)
                /gene="SPAC18G6.02c"
                /note="SPAC18G6.02c, len:960aa; similarity: at N-terminus
                t o SW:HP1_DROVI P29227 heterochromatin protein 1 (HP1)
                (46.4 % identity in 56 aa overlap)"
                /codon_start=1
                /label=chp1
                /product="chromo domain protein; required for centromeric
                silencing and chromosome segregation"
                /protein_id="CAA92382.1"
                /db_xref="GI:1122367"
                /db_xref="SWISS-PROT:Q10103"
                /translation="MVSVKPLPDIDSNEGETDADVYEVEDILADRVNKNGINEYYIKW
                AGYDWYDNTWEPEQNLFGAEKVLKKWKKRKKLIAKGLLEPFDAEDNEAKKMKREKEIL
                RQQRQKRKSELTQLSQKVKEKFKKMRKKPARRIVTIANDEEEEDDQTMDEDAFERKSM
                QGELKERNLTDKTSTLSTSFGETSPDVNPFYLSEWPTVTDSILLSKSLSSDAIPLKNG
                EIKSTMLMPSDSDNSVPGIQNSNNLENTGAFVENANSPQSNTPLSTFRHSSPLSLSPV
                ITSDNDVANSLFFSNSTPLPSSLKIKKEAPKLETHTILVSDNSGSLTKQDILSYFAFI
                KGNIEVFFLKSPKKDKVCNMAYIQFDSIEQAKLAYDKGHPSWHVTLVKGKISTDMEEC
                KVSKSILKTTPSKKANARSVSFTQTTTDTLSESEKFASNVDLDENFDFNVNVTNEDAK
                QLKKSVIGSSWTTVNNDWNSVSKSDQTFENDGASKVVPAGNTTLNSDNSLHHSISESE
                DLSSASTLSDYFRFVLRVGKSLYYAGELSFDISKLKAETEHQQLLRSLVSCKQVDVLR
                FVTSQYLEVFGTCLTKVLSGSLCIRSDVDMTHFKNILNRGNGAGIVLGSNYTLLLFTE
                DNNALMNLYDCQGQSNSPFWMVIFEPLESILVEWSAKNLRPKKPYHKSQSYLSYLLQL
                GHIDLHKIGAFQATQILIVSKQPSPEAEELEDTFREAAIPTFRGLEIPESLFLSQNVF
                VFLNVSLEDDFDQLQFLTLAKRKSCKFFLFGLSLPLKSPNDSHVGTDFKKNNEPLDKL
                TYSQYLRPMFPKGGVVSVTLSALIKTPRLLELISPFLEIKKDSWILILPPSIVDMVKS
                YFVTNNPDKSLLEIQNLLNTLQRYLINPALKNVTLYQDWDIVIDDSADVSLASTLQLY
                QKKNYDKYRRFVLIHELKNELTPVNGLDIVDYDEFKETFMRAIGLK"
misc_feature    complement(5778. .5903)
                /gene="SPAC18G6.02c"
                /note="Ptam match to entry PF00385 chromo, 'chromo'
                (CHRrom atin Organization MOdifier) domain, score 46.20,
                E-value 4. 6e-12"
tRNA            complement(6831. .6912)
                /note="tRNA Val anticodon AAC, Cove score 63.53"
                /product="tRNA-Val"
intron          complement(6867. .6874)
                /note="intron tRNA Val anticodon AAC"
misc_feature    7834. .7889
                /note="intron in 5' utr of transcript for ypt3"
gene            join(7913. .8339,8584. .8664,8719. .8855)
                /gene="ypt3"
                /note="SPAC18G6.03"
CDS             join(7913. .8339,8584. .8664,8719. .8855)
                /gene="ypt3"
                /function="protein transport. probably involved in vesicul
                ar trattic (by similarity)"
                /note="SPAC18G6.03, len:214aa"
                /codon_start=1
                /label=ypt3
                /product="YPT1-related rab subfamily protein"
                /protein_id="CAA92383.1"
                /db_xref="GI:1122368"
                /db_xref="SWISS-PROT:P17610"
                /translation="MCQEDEYDYLFKTVLIGDSGVGKSNLLMRFTRNEFNIESKSTIG
                VEFATRNIVLDNKKIKAQIWDTAGQERYRAITSAYYRGAVGALIVYDITKQSSFDNVG
                RWLKELREHADSNIVIMLVGNKTDLLHLRAVSTEEAQAFAAENNLSFIETSAMDASNV
                EEAFQTVLTEIFRIVSNRSLEAGDDCVHPTAGQTLNIAPIMNDLNKKKSSSQCC"
misc_feature    join(7952. .8339,8584. .8664,8719. .8852)
                /gene="ypt3"
                /note="Match to PF00071 ras, Ras family Score 320.83"
misc_feature    7961. .7984
                /gene="ypt3"
                /note="PS00017 ATP/GTP-binding site motif A"
misc_feature    8340. .8345
                /gene="ypt3"
                /note="splice donor sequence, gtaagt"
misc_feature    8559. .8583
                /gene="ypt3"
                /note "splice branch and acceptor sequence,
                ctaatgttttacag"
misc_feature    8665. .8670
                /gene="ypt3"
                /note="splice donor sequence, gtatgt"
intron          8665. .8718
                /gene="ypt3"
                /note="confirmed by EST"
misc_feature    8703. .8718
                /gene="ypt3"
                /note="splice branch and acceptor sequence,
                tactaatctagccta g"
gene            complement(9285. .10703)
                /gene="shm2"
                /note="SPAC18G6.04c"
CDS             complement(9285. .10703)
                /gene="shm2"
                /note="SPAC18G6.04c, len:472aa; anomoly:possibly spliced
                at 5' end (9288. .10667,10814. .10897)"
                /codon_start=1
                /label=shm2
                /product="serine hydroxymethyltransferase"
                /protein_id="CAA92384.1"
                /db_xref="GI:1122369"
                /db_xref="SWISS-PROT:Q10104"
                /translation="MKCSNKVLTTIKLLKAPLAECDPTVYKILESEKSRQKESIALIA
                SENFTSRAVMDALGSIMQNKYSEGYPGARYYGGNEFIDQAERLCQTRALEAFHLDGEK
                WGVNVQPHSGSPANLQAYQAVMKPHDRLMGLDLPHGGHLSHGFSTPQKAISAVSIYFS
                TMPYNVNKETGTIDYDSLEKAAIQFRPKVIVAGASAYARLVDYKRMKKIIEMCNAYLL
                CDMAHISGLVAAGVIPSPFEYADIVTTTTHKSLRGPRGAMIFYRKGIRSHDKRGNPII
                YELEDKINFSVFPGHQGGPHNHTITALAVALCQAKTPEFYQYQKDVLSNAKAMANAFI
                TRGYKLVSGGTDTHLVLVDLTDKGVDGARVERILELVNISANKNTVPGDKSALIPRGL
                RLGTPACTTRGFDEKDFERVVEFIDEVVSLTKKINEAALKEGKSKFRDFKAYVGDGSK
                FSEIAKLKKEVITWAGKFDFPV"
misc_feature    complement(9396. .9419)
                /gene="shm2"
                /note="PS00017 ATP/GTP-binding site motif A"
misc_feature    complement(9456. .10652)

alignment_scores:
            Quality:  554.50              Length:    634
              Ratio:  1.545                 Gaps:     19
 Percent Similarity:  56.625   Percent Identity: 25.394

alignment_block:
 US-09-932-678-2 x SPAC18G6/rev   ..

 Align seq 1/1  to reverse of: SPAC18G6  from: 1  to: 37621

      26 AlaSerArgThrGlyIleSerAsnMetArgAlaLeuGluAsnAspPhoPh 42
         ||||||   |||:::::::   :::      :::   ::::::::::|||
   30827 GCATCTTCTACAAATGTTCCCGATCGTACTGTAGGTTCCAAAAGTTTTGC 30778

      42 eAsnSerProProArgLysThrValArg.............PheGlyG 54
          :::|||      ::::::      |||                :::   :
   30777 CTCTTCAGTTTCCAAAAATGATGGACGATTAATGCAACAAATGTTACGTG 30728

      54 lyThrValThrGluValLeuLeuLysTyrLysLysGlyGluThrAsnAsp 70
         ::   |||:::::::::::|||           :::|||:::
   30727 CTTTCGTCAATAAAGCTTTGGACGACAAAGCAGAAGGAAATTTTGCCGGT 30678

      71 PheGluLeuLeuLysAsnGlnLeu...LeuAspProAspIleLysAspAs 86
         :::|||   |||:::::::|||:::          |||   |||||
   30677 TATGAAGATCTTAGAAGACAGTTTGCTGCAAAAAGTGATACTAAAGATGC 30628

      86 p...........GlnIleIleAsnTrpLeuLeuGluPheArgSerSerI 99
                    |||:::   |||   |||      :::      ::::
   30627 ACCTTCCTCTCTGCAACTTCAAAACTTATTGTCGGCTTTGACGTGCAATG 30578

      99 leMetTyrLeuThrLysAspPheGluGlnLeuIleSerIleIleLeuArg 115
```

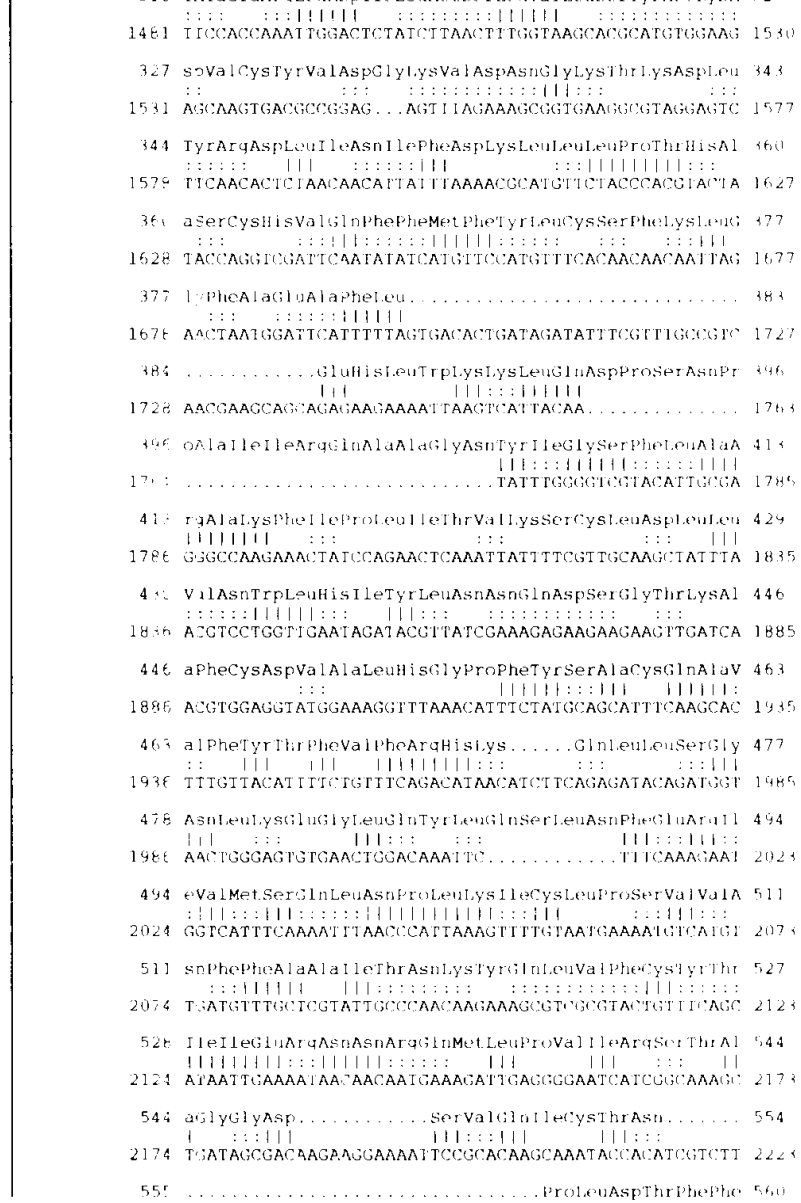


```
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 2672)
  AUTHORS   Yamamoto,R.T., Nogi,Y., Dodd,J.A. and Nomura,M.
  TITLE     RRN3 gene of Saccharomyces cerevisiae encodes an essential RNA
            polymerase I transcription factor which interacts with the
            polymerase independently of DNA template
  JOURNAL   EMBO J. (1996) In press
REFERENCE   2  (bases 1 to 2672)
  AUTHORS   Yamamoto,R.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-1996) Robert T. Yamamoto, Biological Chemistry,
            University of California, Irvine, D240 Medical Science I, Irvine,
            CA, 92717-1700, USA
FEATURES             Location/Qualifiers
     source          1..2672
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
                     /chromosome="XI"
                     /germline
     gene            660..2543
                     /gene="RRN3"
     CDS             660..2543
                     /gene="RRN3"
                     /function="RNA polymerase I-specific transcription factor"
                     /note="corresponds to open reading frame YKL125w from
                     yeast sequencing project (Nature 369:371-8)"
                     /codon_start=1
                     /evidence=experimental
                     /product="Rrn3p"
                     /protein_id="CAA96470.1"
                     /db_xref="GI:1403568"
                     /db_xref="SWISS-PROT:P36070"
                     /translation="MMAFENTSKRPPQDFVAPIDQKKRKVQFSDSTGLVTLQPEEIKD
                     EVFSAAMYSRFVKSALDDLDKNDSTQIGIIANQVALPSKNPERINDKNLNILLDILSS
                     NINRIESSRGTFLIQSIINFEKWWELPPHTLSKYIYFIKILCSSIPKWWQDVSMILVS
                     CFILPIKQTVCHHDMLKYFLRMIPSSMGFIDTYLAKFFPNKNDTRRKLVNYTSNLLKL
                     RGYCSELGFQIWSLLIEKIISIDVELQNELDELDDDVDDDDLEEVDLEDDDDLDDDSG
                     DDDDENCGNSNEELRSGAADGSQSDSEDMDIIEGMDGTEEYNVELTQGIKELSTKLDS
                     ILTLVSTHVEEQVTPESLESGEGVGVFNTLTTLFKTHVLPTYYTRSIQYIMFHVSQQQ
                     LELMDSFLVTLIDISFAVNEAAEKKIKSLQYLGSYIARAKKLSRTQIIFVASYLTSWL
                     NRYVIEREEEVDQRGGMERFKHFYAAFQALCYIFCFRHNIFRDTDGNWECELDKFFQR
                     MVISKFNPLKFCNENVMLMFARIAQQESVAYCFSIIENNNNERLRGIIGKADSDKKEN
                     SAQANTTSSSWSLATRQQFIDLQSYFPYDPLFLKNYKILMKEYYIEWSEASGEYESDG
                     SDD"
BASE COUNT      883 a    451 c    555 g    783 t
ORIGIN


alignment_scores:
           Quality:  424.50             Length:    677
             Ratio:  1.256                Gaps:     18
 Percent Similarity:  49.926   Percent Identity: 23.043

alignment_block:
 US-09-932-678-2 x SCRRN3GEN   ..

 Align seg 1/1   to: SCRRN3GEN  from: 1   to: 2672

     21 ValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLe 37
        |||||||||           :::|||     :::::::|||    |||::
    621 GTAAAGAAATAACTTTTGTGCAAGACAACAGCTGACAAGATGATGGCTTT 670

     37 uGluAsnAspPhePheAsnSerPro......................... 45
        :||||||               :::    |||
    671 TGAGAATACAAGTAAACGACCACCCCAGGACTTTGTAGCTCCTATAGATC 720

     46 ..ProArgLysThrValArgPheGly.........GlyThrValThr... 57
          ::::::  |||:::|||:::           |||    ||||||
    721 AAAAAAAAAGGAAGGTCCAGTTTTCCGATAGTACTGGCCTGGTAACCCTT 770

     57 .................................................. 57

    771 CAACCAGAGGACATAAAAGATGAAGTGTTTTCAGCGGCCATGTATAGCAG 820

     58 .........GluValLeuLeuLysTyrLysLysGlyGluThrAsnAspP 71
                 ::::::|||           |||:::::::::::::::
    821 GTTCGTGAAATCTGCCCTAGATGATCTGGATAAGAATGATTCAACCCAGA 870

     71 heGluLeuLeuLysAsnGlnLeuLeu...........AspProAspIle 83
        ::    ::::::    ||||||:::                :::||||::
    871 TCGGCATAATTGCCAATCAAGTAGCGCTTCCGAGCAAGAATCCTGAAAGA 920

     84 LysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMe 100
          :::|||:::     :::|||   ||||||:::::   ||||||     ::
    921 ATTAATGATAAGAATCTCAATATTCTTTTGGATATCTTATCTAGTAATAT 970

    100 tTyrLeuThrLysAspPheGlu.........GlnLeuIleSerIleIleL 114
        :               ::::::   :::              :::   |||||||||
    971 CAACAGGATAGAATCCTCCAGGGGAACTTTCCTCATACAATCTATCATAA 1020

    114 euArgLeuProTrpLeuAsnArgSerGlnThrValValGluGluTyrLeu 130
                 |||      :::        ::::::::::::|||:::
   1021 ACTTTGAGAAGTGGTGGGAATTGCCTCCACACACTCTAAGTAAATATATA 1070

    131 AlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCy 147
           |||:::      |||    |||:::            :::    :::
   1071 TATTTCATCAAAATCCTTTGCTCGAGTATACCCAAATGGTGGCAAGACGT 1120

    147 sLeuSerMetIleAlaSerHisPheValProProArgValIleIleLysG 164
              ::::::::::|||    |||:::
   1121 TTCGATGATACTGGTATCTTGTTTTATT...................... 1148

    164 luGlyAspValAspValSerAspSerAspAspGluAspAspAsnLeuPro 180
                                                    ||||||
   1149 ..........................................TTACCA 1154

    181 AlaAsnPheAspThrCysHisArg...AlaLeuGlnIleIleAlaArgTy 196
           :::        :::||||||:::        |||:::    :::    |||
   1155 ATCAAACAGACAGTGTGTCATCATGATATGCTTAAATATTTCTTAAGGAT 1204

    196 rValProSerThrProTrpPheLeuMetProIleLeuValGluLysPheP 213
          :::||||||:::          |||:::        |||:::::    ||||
   1205 GATTCCCTCCTCGATGGGATTCATAGATACATATTTGGCCAAATTTTTCC 1254

    213 roPheValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeu 229
         ||        :::    :::::::|||    |||       |||:::    ||||||
   1255 CAAATAAAAATGATACTCGGAGAAAGTTGGTTAATTACACTTCAAATTTG 1304

    230 LeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLe 246
        |||::::::       |||             |||       :::|||     :::||
   1305 CTGAAATTGAGAGGCTACTGCTCTGAGCTAGGATTCCAAATTTGGTCCTT 1354

    246 uIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGln.... 261
        |:::||||||||||::::::::::::::||||||:::     :::::::::
   1355 ACTGATAGAAAACATTATTTCGATTGATGTCGAGTTACAAAATGAATTAG 1404

    262 ....................GlyIleGluAspAlaGluGluThrAlaThr 271
                                 :::|||    ||||||:::
   1405 ATGAATTAGATGATGATGTAGATGACGATGATCTAGAGGAAGTTGATCTA 1454

    272 GlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGl 288
        :.:                  |||:::    :::|||                |||::
   1455 GAGGATGACGACGATTTAGATGATGATTCAGGA...........GACGA 1492

    288 uAspGluGlu...........ThrGluHisGluThrLysAlaGlyProG 301
        :|||:::|||              ::::::::::|||    ::::::|||
   1493 TGATGATGAAAATTGTGGGAACAGTAATGAAGAGTTAAGGTCTGGAGCAG 1542

    301 lu.............................................. 301

   1543 CCGACGGTAGTCAAAGTGATAGTGAAGATATGGACATCATTGAAGGTATG 1592

    302 .......................ArgLeuAspGlnMetValHisProVa 310
                               :::|||   |||    :::       ::
   1593 GACGGTACTGAAGAATATAATGTGGAGTTGACGCAGGGAATCAAAGAACT 1642
```

```
 310 lAlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysA 327
     ::::   :::||||||    :::::::::|||||    :::::::::::::
1643 TTCCACCAAATTGGACTCTATCTTAACTTTGGTAAGCACGCATGTGGAAG 1692

 327 spValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeu 343
     ::           :::    ::::::::::::|||:::          :::
1693 AGCAAGTGACGCCGGAG...AGTTTAGAAAGCGGTGAAGGCGTAGGAGTC 1739

 344 TyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAl 360
     ::::::    |||    ::::::|||             :::||||||||:::
1740 TTCAACACTCTAACAACATTATTTAAAACGCATGTTCTACCCACGTACTA 1789

 360 aSerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuG 377
       :::        :::|||::::::||||||::::::    :::    :::|||
1790 TACCAGGTCGATTCAATATATCATGTTCCATGTTTCACAACAACAATTAG 1839

 377 lyPheAlaGluAlaPheLeu.............................. 383
        :::    ::::::||||||
1840 AACTAATGGATTCATTTTTAGTGACACTGATAGATATTTCGTTTGCCGTC 1889

 384 .............GluHisLeuTrpLysLysLeuGlnAspProSerAsnPr 396
                  |||          |||::::||||||
1890 AACGAAGCAGCAGAGAAGAAAATTAAGTCATTACAA.............. 1925

 396 oAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaA 413
                                  |||:::||||||::::::||||
1926 ...........................TATTTGGGGTCGTACATTGCGA 1947

 413 rgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeu 429
     ||||||||     :::              :::             :::    |||
1948 GGGCCAAGAAACTATCCAGAACTCAAATTATTTTCGTTGCAAGCTATTTA 1997

 430 ValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAl 446
     ::::::||||||:::     |||:::    ::::::::::::     :::
1998 ACGTCCTGGTTGAATAGATACGTTATCGAAAGAGAAGAAGAAGTTGATCA 2047

 446 aPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaV 463
               :::                    ||||||:::|||    ||||||:
2048 ACGTGGAGGTATGGAAAGGTTTAAACATTTCTATGCAGCATTTCAAGCAC 2097

 463 alPheTyrThrPheValPheArgHisLys......GlnLeuLeuSerGly 477
     ::     |||    |||    |||||||||:::          :::     :::|||
2098 TTTGTTACATTTTCTGTTTCAGACATAACATCTTCAGAGATACAGATGGT 2147

 478 AsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIl 494
     |||    :::          |||:::    :::                |||:::|||::
2148 AACTGGGAGTCTGAACTGGACAAATTC............TTTCAAAGAAT 2185

 494 eValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValA 511
     :|||:::|||::::::|||||||||||||:::|||        :::|||:::
2186 GGTCATTTCAAAATTTAACCCATTAAAGTTTTGTAATGAAAATGTCATGT 2235

 511 snPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThr 527
       :::||||||    |||:::::::::    ::::::::::::|||::::::
2236 TGATGTTTGCTCGTATTGCCCAACAAGAAAGCCGTCGCCGTACTGTTTCAGC 2285

 528 IleIleGluArgAsnAsnArgGlnMetLeuProValIleArgSerThrAl 544
     |||||||||:::||||||:::::    |||        |||    :::    ||
2286 ATAATTGAAAATAACAACAATGAAAGATTGAGGGGAATCATCGGCAAAGC 2335

 544 aGlyGlyAsp............SerValGlnIleCysThrAsn....... 554
     |     :::|||                |||:::|||        |||:::
2336 TGATAGCGACAAGAAGGAAAATTCCGCACAAGCAAATACCACATCGTCTT 2385

 555 ..................................ProLeuAspThrPhePhe 560
                                        |||::::::::::|||
2386 CTTGGTCTTTAGCCACTCGACAACAATTTATTGACTTACAGAGTTACTTC 2435

 561 ProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIl 577
     |||:::||||||          ||||||:::    |||     ::::::
2436 CCATATGATCCTTTGTTTCTGAAAAACTATAAAATTTTAATGAAGGAATA 2485

 577 eTyrGlnValTrpGluAspMetSerAlaGlu 587
```

```
     |||          |||::::::    |||:::|||
2486 CTACATAGAGTGGAGTGAAGCAAGCGGGGAA 2516

seq_name: gb_pl:SCYKL125W

seq_documentation_block:
LOCUS       SCYKL125W               2843 bp    DNA     linear   PLN 11-AUG-1997
DEFINITION  S.cerevisiae chromosome XI reading frame ORF YKL125w.
ACCESSION   Z28125 Y13137
VERSION     Z28125.1  GI:486210
KEYWORDS    .
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 2843)
  AUTHORS   Ramezani Rad,M., Xu,G., Kirchrath,L., Fritz,C., Keuchel,H. and
            Hollenberg,C.P.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2843)
  AUTHORS   MIPS.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-1994) Data collected by MIPS on behalf of the
            European yeast chromosome XI sequencing project. MIPS at the
            Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
            Martinsried, FRG; E-mail: Mewes@ehpmic.mips.biochem.mpg.de
FEATURES             Location/Qualifiers
     source          1..2843
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
                     /chromosome="XI"
     gene            498..2381
                     /gene="RRN3"
     CDS             498..2381
                     /gene="RRN3"
                     /note="ORF YKL125w"
                     /codon_start=1
                     /protein_id="CAA81966.1"
                     /db_xref="GI:486211"
                     /db_xref="SWISS-PROT:P36070"
                     /translation="MMAFENTSKRPPQDFVAPIDQKKRKVQFSDSTGLVTLQPEEIKD
                     EVFSAAMYSRFVKSALDDLDKNDSTQIGIIANQVALPSKNPERINDKNLNILLDILSS
                     NINRIESSRGTFLIQSIINFEKWWELPPHTLSKYIYFIKILCSSIPKWWQDVSMILVS
                     CFILPIKQTVCHHDMLKYFLRMIPSSMGFIDTYLAKFFPNKNDIRRKLVNYTSNLLKL
                     RGYCSELGFQIWSLLIEKIISIDVELQNELDELDDDVDDDDLEEVDLEDDDDLDDDSG
                     DDDDENCGNSNEELRSGAADGSQSDSEDMDIIEGMDGTEEYNVELTQGIKELSTKLDS
                     ILTLVSTHVEEQVTPESLESGEGVGVFNTLTTLFKTHVLPTYYTRSIQYIMFHVSQQQ
                     LELMDSFLVTLIDISFAVNEAAEKKIKSLQYLGSYIARAKKLSRTQIIFVASYLTSWL
                     NRYVIEREEEVDQRGGMERFKHFYAAFQALCYIFCFRHNIFRDTDGNWECELDKFFQR
                     MVISKFNPLKFCNENVMLMFARIAQQESVAYCFSIIENNNNERLRGIIGKADSDKKEN
                     SAQANTTSSSWSLATRQQFIDLQSYFPYDPLFLKNYKILMKEYYIEWSEASGEYESDG
                     SDD"
BASE COUNT      935 a    482 c    565 g    861 t
ORIGIN


alignment_scores:
         Quality:  424.50              Length:    677
           Ratio:  1.256                 Gaps:     18
 Percent Similarity:  49.926   Percent Identity:  23.043

alignment_block:
 US-09-932-678-2 x SCYKL125W    ..

 Align seq 1/1  to: SCYKL125W  from: 1  to: 2843

   21 ValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnMetArgAlaLe 37
      |||||||||             :::|||       ::::::|||    |||::
  459 GTAAAGAAATAACTTTTGTGCAAGACAACAGCTGACAAGATGATGGCTTT 508

   37 uGlnAsnAspPhePheAsnSerPro......................... 45
      :||||||          :::    |||
  509 TCAGAATACAAGTAAACCACCACCCAGGACTTTGTAGCTCCTATAGATC 558
```


46 ..ProArgLysThrValArgPheGly.....GlyThrValThr... 57
559 AAAAAAAAAGGAAGGTCCAGTTTTCCGATAGTACTGGCCTGGTAACCCTT 608
57 57
609 CAACCAGAGGAGATAAAAGATGAAGTGTTTTCAGCGGCCATGTATAGCAG 658
58 .....GluValLeuLeuLysTyrLysLysGlyGluThrAsnAspP 71
659 GTTCGTGAAATCTGCCCTAGATGATCTGGATAAGAATGATTCAACCCAGA 708
71 heGluLeuLeuLysAsnGlnLeuLeu.....AspProAspIle 83
709 TCGGCATAATTGCCAATCAAGTAGCGCTTCCGAGCAAGAATCCTGAAAGA 758
84 LysAspAspGlnIleIleAsnTrpLeuLeuGluPheArgSerSerIleMe 100
759 ATTAATGATAAGAATCTCAATATTCTTTTGGATATCTTATCTAGTAATAT 808
100 tTyrLeuThrLysAspPheGlu.....GlnLeuIleSerIleIleL 114
809 CAACAGGATAGAATCCTCCAGGGGAACTTTCCTCATACAATCTATCATAA 858
114 euArgLeuProTrpLeuAsnArgSerGlnThrValValGluGluTyrLeu 130
859 ACTTTGAGAAGTGGTGGGAATTGCCTCCACACACTCTAAGTAAATATATA 908
131 AlaPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProCy 147
909 TATTTCATCAAAATCCTTTGCTCGAGTATACCCAAATGGTGGCAAGACGT 958
147 sLeuSerMetIleAlaSerHisPheValProProArgValIleIleLysG 164
959 TTCGATGATACTGCTATCTTGTTTTATT..... 986
164 luGlyAspValAspValSerAspSerAspAspGluAspAspAsnLeuPro 180
987 .....TTACCA 992
181 AlaAsnPheAspThrCysHisArg...AlaLeuGlnIleIleAlaArgTy 196
993 ATCAAACAGACAGTGTGTCATCATGATATGCTTAAATATTTCTTAAGGAT 1042
196 rValProSerThrProTrpPheLeuMetProIleLeuValGluLysPheP 213
1043 GATTCCCTCCTCGATGGGATTCATAGATACATATTTGGCCAAATTTTTCC 1092
213 roPheValArgLysSerGluArgThrLeuGluCysTyrValHisAsnLeu 229
1093 CAAATAAAAATGATACTCGGAGAAAGTTGGTTAATTACACTTCAAATTTG 1142
230 LeuArgIleSerValTyrPheProThrLeuArgHisGluIleLeuGluLe 246
1143 CTGAAATTGAGAGGCTACTGCTCTGAGCTAGGATTCCAAATTTGGTCCTT 1192
246 uIleIleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGln.... 261
1193 ACTGATAGAAAAGATTATTTCGATTGATGTCGAGTTACAAAATGAATTAG 1242
262 .....GlyIleGluAspAlaGluGluThrAlaThr 271
1243 ATGAATTAGATGATGATGTAGATGACGATGATCTAGAGGAAGTTGATCTA 1292
272 GlnThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspGl 288
1293 GAGGATCACGACGATTTAGATGATGATTCAGGA.....GACGA 1330
288 uAspGluGlu.....ThrGluHisGluThrLysAlaGlyProG 301
1331 TGATGATGAAAATTGTGGGAACAGTAATGAAGAGTTAAGGTCTGGAGCAG 1380


301 lu..... 301
1381 CCGACGGTAGTCAAAGTGATAGTGAAGATATGGACATCATTGAAGGTATG 1430
302 .....ArgLeuAspGlnMetValHisProVa 316
1431 GACGGTACTGAAGAATATAATGTGGAGTTGACGCAGGGAATCAAAGAACT 1480
317 lAlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysA 327
1481 TTCCACCAAATTGGACTCTATCTTAACTTTGGTAAGCACGCATGTGGAAG 1530
327 spValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeu 343
1531 AGCAAGTGACGCCGGAG...AGTTTAGAAAGCGGTGAAGGCGTAGGAGTC 1577
344 TyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAl 360
1578 TTCAACACTCTAACAACATTATTTAAAACGCATGTTCTACCCACGTACTA 1627
361 aSerCysHisValGlnPhePheMetPheTyrLeuCysSerPheLysLeuG 377
1628 TACCAGGTCGATTCAATATATCATGTTCCATGTTTCACAACAACAATTAG 1677
377 lyPheAlaGluAlaPheLeu..... 383
1678 AACTAATGGATTCATTTTTAGTGACACTGATAGATATTTCGTTTGCCGTC 1727
384 .....GluHisLeuTrpLysLysLeuGlnAspProSerAsnPr 396
1728 AACGAAGCAGCAGAGAAGAAAATTAAGTCATTACAA..... 1763
396 oAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerPheLeuAlaA 413
1764 .....TATTTGGGGTCGTACATTGCGA 1789
413 rgAlaLysPheIleProLeuIleThrValLysSerCysLeuAspLeuLeu 429
1786 GGGCCAAGAAACTATCCAGAACTCAAATTATTTTCGTTGCAAGCTATTTA 1835
430 ValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAl 446
1836 ACGTCCTGGTTGAATAGATACGTTATCGAAAGAGAAGAAGAAGTTGATCA 1885
446 aPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaV 463
1886 ACGTGGAGGTATGGAAAGGTTTAAACATTTCTATGCAGCATTTCAAGCAC 1935
463 alPheTyrThrPheValPheArgHisLys.....GlnLeuLeuSerGly 477
1936 TTTGTTACATTTTCTGTTTCAGACATAACATCTTCAGAGATACAGATGGT 1985
478 AsnLeuLysGluGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIl 494
1986 AACTGGGAGTGTGAACTGGACAAAATTC.....TTTCAAAGAAT 2024
494 eValMetSerGlnLeuAsnProLeuLysIleCysLeuProSerValValA 511
2024 GGTCATTTCAAAATTTAACCCATTAAAGTTTTGTAATGAAAATGTCATGT 2073
511 snPhePheAlaAlaIleThrAsnLysTyrGlnLeuValPheCysTyrThr 527
2074 TGATGTTTGCTCGTATTGCCCAACAAGAAAGCGTCGCGTACTGTTTCAGC 2123
528 IleIleGluArgAsnAsnArgGlnMetLeuProValIleArgSerThrAl 544
2124 ATAATTGAAAATAACAACAATGAAAGATTGAGGGAATCATCGGCAAAGC 2173
544 aGlyGlyAsp.....SerValGlnIleCysThrAsn..... 554
2174 TGATAGCGACAAGAAGGAAAATTCCGCACAAGCAAATACCACATCGTCTT 2223
555 .....ProLeuAspThrPhePhe 560

```
                                                 |||:::::::::::|||
    2224 CTTGGTCTTTAGCCACTCGACAACAATTTATTGACTTACAGAGTTACTTC 2273

     561 ProPheAspProCysValLeuLysArgSerLysLysPheIleAspProIl 577
         |||:::|||||||      |||||||:::  |||   ::::::
    2274 CCATATGATCCTTTGTTTCTGAAAAACTATAAAATTTTAATGAAGCAATA 2323

     577 oTyrGlnValTrpGluAspMetSerAlaGlu 587
         |||      |||:::::::   |||:::|||
    2324 CTACATAGAGTGGAGTGAACCAAGCGGGGAA 2354

Seq name: gb pr:AC106788

Seq documentation block:
LOCUS       AC106788               170611 bp    DNA     linear   PRI 03-FEB-2002
DEFINITION  Homo sapiens chromosome 16 clone RP11 368J21, complete sequence.
ACCESSION   AC106788
VERSION     AC106788.2  GI:18482004
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170611)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170611)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (12 JAN 2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 170611)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03 FEB 2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Feb 4, 2002 this sequence version replaced gi:18139338.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base by base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
                 .       .     .
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base by base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
                 .       .     .
FEATURES             Location/Qualifiers
     source          1..170611
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "16"
                     /clone "RP11 368J21"
BASE COUNT    47955 a   38538 c   36807 g   45411 t
ORIGIN


alignment scores:
            Quality:   498.00          Length:     396
              Ratio:     4.790            Gaps:       1
 Percent Similarity:  26.515   Percent Identity: 26.010
```

```
alignment block:
 US-09-932-678-2 x AC106788  ..

 Align seq 1/1  to: AC106788  from: 1  to: 170611

     477 GlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspPr 893
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  154053 GGATCGCAGAGGCATTTTTGGAACATCTTTGGAAAAAATTGCAGGATCC 154102

     893 oSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerP 410
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  154103 AAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCT 154152

     410 heLeuAlaArgAlaLysPheIleProLeuIle..................... 420
         ||||||||||||||||||||||||||||||:::
  154153 TTTTGGCAAGAGCTAAATTTATTCCTCTTATGTAAGTAGCCTAATTTTCC 154202

     420 ..................................................... 420

  154203 GAGTACTTTTTAATATCATGCTTAAAAAGAGTATAGCATTGTCTCAAGT 154252

     420 ..................................................... 420

  154253 CAGAAATATCTCCCATATGTTTTTGGCATGTTTTTAAAGTGAATAAAATT 154302

     420 ..................................................... 420

  154303 CCTACTCTGTGTAAGATGTTTATATTTCTAAGTGGTGATTTAGAATAAA 154352

     420 ..................................................... 420

  154353 GTGTCTGCTTTTTTATATATAAAACCCTGTATGTAAGGCTTTTGTCATCT 154402

     420 ..................................................... 420

  154403 CTTTTGTGTGGTTGCACTTAAAGATCCATTTGTTTTGTGGATAGAGGACA 154452

     420 ..................................................... 420

  154453 GTGTTGTATACTGTTTTGATTCTTTTTGTAGGTTTGTCATTTTTTCATTT 154502

     420 ..................................................... 420

  154503 CTATTCCAAATCTATTGTATCTGTTAAAGCTGAAGAAAAACCCTTTTAAA 154552

     420 ..................................................... 420

  154553 GGTAATAGACCTATCTAGGAGGCCAGTTTCTTTCAGTGGCCCATGAAGAT 154602

     420 ..................................................... 420

  154603 ATCTTTGGACAAGGATGCTGTTGAATCCCTTCCCCAAGAACAAAATTATT 154652

     420 ..................................................... 420

  154653 CACCATAGGACTTGACTGGATGCATCAGGCAATACTGAAGTCCACCAGAC 154702

     420 ..................................................... 420

  154703 TGTCTTTCTCTCTTGAGACGTGTTGGTGAACATGTCCTATTTGCCCAATCAC 154752

     420 ..................................................... 420

  154753 CCTAAGAGGGGTGCCTTTGAGATGGTTAGGAGAACCTGCTTTCCATCCCT 154802

     420 ..................................................... 420

  154803 TGGGACGTTCTTAGGGGCTCACCTGTTCCTAGAAGGTCAGAGCTACTCTG 154852

     420 ..................................................... 420

  154853 CCTTGTAATTGGAAGGTTGTCTTCCTACGCACCCATCCTTATCCTTCCTT 154902
```

```
   420 ..................................................  420

154903 TCTTTGCTTTTCCTCTGTACCCATGGGTATTATTIAAAGAAACCTATGAA 154952

   420 ..................................................  420

154953 CTTACTTAGCATGGTTTGTAATGAAAGGCAGTTGTGTGTTTTTATGTTAT 155002

   420 ..................................................  420

155003 TCTGGTTTTTTTATGAAGTGTAAAGTTGACTTGAATTTTTCTTTTCTCTA 155052

   421 ..ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuHisIleTyr 436
         |||||||||    ||||||||||||||||||||||||||||||||||||
155053 GTACTGTAAAACCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATAC 155102

   437 LeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHi 453
         ||||||||||||||||||||||||||||||||||:::||||||||||||
155103 CTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCAATGTTGCTCTCCA 155152

   453 sGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheA 470
         ||||||||||||||||||||||||||||||||||||||||||||||||
155153 TGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTA 155202

   470 rgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGly 482
         |||||||||||||||||||||||||||||||||||||
155203 GACACAAGCAGCTTTTGAGCGGAAACCTGAAAGAAGGT 155240

seq_name: gb_htg:AC007615

seq_documentation_block:
LOCUS       AC007615              175691 bp    DNA     linear   HTG 03-JUL-2001
DEFINITION  Homo sapiens chromosome 16 clone RP11-528K16, WORKING DRAFT
            SEQUENCE, 1 ordered pieces.
ACCESSION   AC007615
VERSION     AC007615.6  GI:14589428
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175691)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175691)
  AUTHORS   Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
            Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
            Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
            Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
            McMurry,K., Han,C. and Deaven,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint
            Genome Institute, Los Alamos National Laboratory, MS M888, Los
            Alamos, NM 87545, USA
COMMENT     On Jul 3, 2001 this sequence version replaced gi:13928651.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
            ----------------------.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1   175691: contig of 175691 bp in length.
FEATURES             Location/Qualifiers
     source          1..175691
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-528K16"
BASE COUNT    42392 a  42125 c  45277 g  45896 t      1 others
ORIGIN


alignment_scores:
            Quality:  398.00              Length:     396
              Ratio:    4.790               Gaps:       1
 Percent Similarity:   26.515   Percent Identity: 26.263

alignment_block:
 US-09-932-678-2 x AC007615   ..

 Align seq 1/1  to: AC007615  from: 1  to: 175691

    377 GlyPheAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspPr 393
        |||||||||||||||||||||||    |||||||||||||||||||||||
  30005 GGATTCGCAGAGGCATTTTTGGAACCTCTTTGGAAAAAATTGCAGGATCC 30054

    393 oSerAsnProAlaIleIleArgGlnAlaAlaGlyAsnTyrIleGlySerP 410
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  30055 AAGTAATCCTGCCATCATCAGGCAGGCTGCTGGAAATTATATTGGAAGCT 30104

    410 heLeuAlaArgAlaLysPheIleProLeuIle..................... 420
        |||||||||||||||||||||||||||||:::
  30105 TTTTGGCAAGAGCTAAATTTATTCCTCTTATGTAAGTAGCCTAATTTTCT 30154

    420 ..................................................  420

  30155 GAGTACTTTTTAATATCATGCTTTAAAAAGAGTATAGCATTGTCTCAAGT 30204

    420 ..................................................  420

  30205 CAGAAATATCTCCCATATGTTTTTGGCATGTTTTTAAAGTGAATAAAATT 30254

    420 ..................................................  420

  30255 CCTACTCTGTGCAAGATGTTTATATTTCTAAGTGGTGATTTTAGAATAAA 30304

    420 ..................................................  420

  30305 GTGTCTCCTTTTCTATATATAAAACCCTGTATGTAAGGCTTTTGTCATCT 30354

    420 ..................................................  420

  30355 CTTTGTGTGGTTGCACTTAAAGATCCATTTGTTTTGTGCATAGAGGACA 30404

    420 ..................................................  420

  30405 GTGTTGTATACTGTTTTGATTCTTTTTGTAGGTTTGTCATTTTTTCATTT 30454

    420 ..................................................  420

  30455 GCATTCCAAATCTATTGTATCTGTTAAAGCTGAAGAAAAACCCTTTTAAA 30504
```

```
  420 ..................................................  420

30505 GGTAATAGACCTATCTAGGAGGCCAGTTTCTTGCAGTGGCCCATGAAGAT 30554

  420 ..................................................  420

30555 ATCTTTGGACAAGGATGCTGTTGAAACCCTTCCCCAAGAACAAAATTATT 30604

  420 ..................................................  420

30605 CACCATAGGACTTGACTGGATGCATCAGGGAATACTGAAGTCCACCACAC 30654

  420 ..................................................  420

30655 TGTCTTTCTCTTGAGAGGTGTTGGTGAACATGTCCTATTTGGCCAATCAC 30704

  420 ..................................................  420

30705 CCTAAGAGGGGTGCCTTTGAGATGGTTAGGAGAACCTGCTTTCCATCCCT 30754

  420 ..................................................  420

30755 TGGGACTTTCTTAGGGGCTCACCTGTTCCTAGAAGGTCAGAGCTACTCTG 30804

  420 ..................................................  420

30805 CCTTGTAATTGGAAGGTTGTCTTCCTACGCACTCATCCTTATCCTTCCTT 30854

  420 ..................................................  420

30855 TCTTTGCTTTTCCTCTGTACCCATGGGTATTATTTAAAGAAACCTATGAA 30904

  420 ..................................................  420

30905 CTTACTTAGCATGGTTTGTAATGAAAGGCAGTTGTGTGTTTTTATGTTAT 30954

  420 ..................................................  420

30955 TCTGGTTTTTTTATGAACTGTAAAGTTGATTTGAATTTTTCTTTCCTCTA 31004

  421 ..ThrValLysSerCysLeuAspLeuLeuValAsnTrpLeuIleIleTyr 436
        ||||||||||||||||||||||||||||||||||||||||||||||||||
31005 GTACTGTAAAATCATGCCTAGATCTTTTGGTTAACTGGCTGCACATATAC 31054

  437 LeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHi 453
      ||||||||||||||||||||||||||||||||||||||||||||||||||
31055 CTTAATAACCAGGATTCGGGAACAAAGGCATTCTGCGATGTTGCTCTCCA 31104

  453 sGlyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValPheA 470
      ||||||||||||||||||||||||||||||||||||||||||||||||||
31105 TGGACCATTTTACTCAGCCTGCCAAGCTGTGTTCTACACCTTTGTTTTTA 31154

  470 rgHisLysGlnLeuLeuSerGlyAsnLeuLysGluGly 482
      |||||||||||||||||||||||||||||||||||||
31155 GACACAAGCAGCTTTGAGCGGAAACCTGAAAGAAGGT 31192
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 02:39:01 ; Search time 242.45 Seconds
(without alignments)
14644.583 Million cell updates/sec

Title: US-09-932-678-1
Perfect score: 2068
Sequence: 1 acagaggctgtggctggaaq.....tgacatttgggattcccccat 2068

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2018.4	97.6	2040	21	AAA98384	Human RNA polymera
2	1228.4	59.4	1418	21	AAC76937	Human ORFX ORF2492
3	1010.4	48.9	1498	22	AAS44584	Human full-length
4	894	43.2	1423	22	AAS01562	Human secretory mo
5	639.4	30.9	2410	23	AAS92255	DNA encoding novel
6	568.6	27.5	776	23	AAS92253	DNA encoding novel
7	550.8	26.6	1260	23	AAS87643	DNA encoding novel
c 8	477.8	23.1	655	23	AAS80955	DNA encoding novel
9	466.2	22.5	3169	23	AAS87641	DNA encoding novel
10	420.8	20.3	1461	22	ABA83038	Human transcriptio
11	420	20.3	437	20	AAX51663	Human secreted pro
12	398	19.2	560	23	AAS91991	DNA encoding novel
13	390.6	18.9	3175	23	AAS92254	DNA encoding novel
c 14	387	18.7	585	23	AAS92248	DNA encoding novel
15	315	15.2	1017	23	AAS87636	DNA encoding novel
16	292	14.1	341	20	AAV90248	EST clone D1349.
17	286.6	13.9	498	23	AAS92247	DNA encoding novel
18	282.2	13.6	396	23	AAS80954	DNA encoding novel
19	282.2	13.6	396	23	AAS87635	DNA encoding novel
20	282.2	13.6	396	23	AAS91986	DNA encoding novel
21	282.2	13.6	396	23	AAS92246	DNA encoding novel
22	279	13.5	2991	23	AAS73660	DNA encoding novel
23	279	13.5	2991	23	AAS91992	DNA encoding novel
24	247.4	12.0	349	21	AAC77533	Human ORFX ORF3088
25	233.4	11.3	667	23	AAS92251	DNA encoding novel
26	230	11.1	2103	21	AAC77798	Human cancer assoc
27	208	10.1	34269	22	AAK68677	Human immune/haema
28	208	10.1	34269	22	AAK85168	Human immune/haema
29	203.6	9.8	213	21	AAC19091	Human secreted pro
30	192	9.3	17803	22	AAK68676	Human immune/haema
31	188.8	9.1	9259	22	AAK68674	Human immune/haema
32	188.8	9.1	23815	22	AAK68678	Human immune/haema
33	188.8	9.1	23815	22	AAK85169	Human immune/haema
34	163.4	7.9	627	23	AAS87640	DNA encoding novel
35	144.2	7.0	3410	22	AAK85167	Human immune/haema
c 36	139.4	6.7	1326	23	AAS91907	DNA encoding novel
37	139.4	6.7	3042	23	AAS70319	DNA encoding novel
c 38	138.6	6.7	807	23	AAS92252	DNA encoding novel
39	128.2	6.2	421	21	AAC10753	Human secreted pro
40	125.8	6.1	255	21	AAC10754	Human secreted pro
41	118.2	5.7	330	16	AAT20548	Human gene signatu
42	106.6	5.2	414	22	ABA46215	Human breast cell
43	106.6	5.2	414	22	ABA56760	Human foetal liver
44	106.6	5.2	414	22	ABA26382	Probe #4848 for ge
45	106.6	5.2	414	22	AAK04888	Human brain expres

ALIGNMENTS

RESULT 1
AAA98384
ID AAA98384 standard; cDNA; 2040 BP.
XX
AC AAA98384;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human RNA polymerase I transcription factor TIF 1A cDNA.
XX
KW RNA polymerase I transcription factor TIF 1A; antitumor; treatment;
KW antiproliferative; cell proliferation; cancer; tissue regeneration; ss.
XX
OS Homo sapiens.
XX
PN WO200055316-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-DE00767.
XX
PR 17-MAR-1999; 99DE-1011992.
XX
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI Grummt I, Vingron M;
XX
DR WPI; 2000-587527/55.
DR P-PSDB; AAB10936.
XX
PT New DNA encoding the transcription factor TIF-1A, useful for preventing
PT or treating diseases associated with abnormal cell proliferation.

```
PT   particularly tumors
XX
PS   Claim 1; Fig 2; 38pp; German.
XX
CC   This invention describes a novel DNA sequence (I) that encodes the RNA
CC   polymerase I transcription factor TIF-IA which has antitumor,
CC   antiproliferative and proliferation-inducing activity. The invention also
CC   describes (1) DNA (Ia) encoding a protein (II) with the biological
CC   activity of TIF-IA; (2) a ribozyme (R) corresponding to (I) or (Ia) and
CC   able to bind specifically to, and cleave, its transcribed RNA so as to
CC   reduce or inhibit synthesis of the corresponding protein; (3) an
CC   antisense RNA (AS) with binding properties similar to R; (4) an
CC   expression vector that contains (I), (Ia) or sequences that encode R or
CC   AS; (5) host cells containing the vectors of (4); (6) TIF-IA or (II)
CC   encoded by (I) or (Ia); (7) preparation of TIF-IA or (II) by culturing
CC   cells of (6); (8) ligands that bind to TIF-IA or (II); (9) antagonists
CC   that weaken or block the activity of TIF-IA or (II); (10) a diagnostic
CC   method for detecting abnormal TIF-IA expression; and (11) kit for
CC   carrying out the method in (10). (I), and similar sequences that encode
CC   proteins with equivalent activity, expression vectors containing them, or
CC   their expression products are used to treat or prevent disorders
CC   associated with reduced cellular proliferation, to stimulate cellular
CC   proliferation, and to promote tissue regeneration, e.g. after injury or
CC   radiation therapy. Ribozymes, antisense sequences directed against (I),
CC   also ligands and antagonists of TIF IA, are used to treat or prevent
CC   disorders associated with excessive cellular proliferation and to inhibit
CC   proliferation, especially in treatment of cancers. (I) and specific
CC   ligands for TIF-IA (particularly antibodies (Ab)) are also useful for
CC   diagnosis of altered TIF-IA expression by (in)direct determination of the
CC   concentration, length and/or sequence of TIF-IA or its mRNA, e.g. for
CC   detecting mutations. Ab can also be used for immunoprecipitation of
CC   TIF IA and for isolation of related sequences from cDNA expression
CC   libraries. (I) allows recombinant production of TIF-IA in sufficient
CC   quantities for therapeutic use.
XX
SQ   Sequence 2040 BP; 563 A; 433 C; 474 G; 568 T; 2 other;

  Query Match             97.6%;  Score 2018.4;  DB 21;  Length 2040;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2019;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         47 gcggtcgcgttagttcggcccaatggcggcaccgctgcttcacacgcgttgccgggaga 106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         18 gcggtcgcgttagttcggcccaatggcggcaccgctgcttcacacgcgttgccgggaga 77

Qy        107 tgcggccgcttcgtcctctgcagttaagaagctgggcgcgtcgaggactgggatttcaaa 166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 tgcggccgcttcgtcctctgcagttaagaagctgggcgcgtcgaggactgggatttcaaa 137

Qy        167 tatgcgtgcattagagaatgactttttcaattctcccccaagaaaaactgttcggtttgg 226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 tatgcgtgcattagagaatgactttttcaattctcccccaagaaaaactgttcggtttgg 197

Qy        227 tggaactgtgacagaagtcttgctgaagtacaaaaagggtgaaacaaatgactttgagtt 286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 tggaactgtgacagaagtcttgctgaagtacaaaaagggtgaaacaaatgactttgagtt 257

Qy        287 gtttgaagaaccagctgttagatccagacataaaggatgaccagatcatcaactggctgct 346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 gtttgaagaaccagctgttagatccagacataaaggatgaccagatcatcaactggctgct 317

Qy        347 agaattccgttcttctatcatgtacttgacaaaagactttgagcaacttatcagtattat 406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 agaattccgttcttctatcatgtacttgacaaaagactttgagcaacttatcagtattat 377

Qy        407 attaagattgccttggttgaatagaagtcaaacagtagtggaagagtatttggcttttct 466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 attaagattgccttggttgaatagaagtcaaacagtagtggaagagtatttggcttttct 437

Qy        467 tggtaatcttgtatcagcacagactgttttcctcagaccgtgtctcagcatgattgcttc 526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 tggtaatcttgtatcagcacagactgttttcctcagaccgtgtctcagcatgattgcttc 497

Qy        527 ccattttgtgcctccccgagtgatcattaaggaaggcgatgtagatgtttcagattctga 586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 ccattttgtgcctccccgagtgatcattaaggaaggcgatgtagatgtttcagattctga 557

Qy        587 tgatgaagatgataatcttcctgcaaattttgacacatgtcacagagccttgcaaataat 646
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 tgatgaagatgataatcttcctgcaaattttgacacatgtcacagagccttgcaaataat 617

Qy        647 agcaagatatgtaccatcgacaccgtggtttctcatgccaatactggtggaaaaatttcc 706
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 agcaagatatgtaccatcgacaccgtggtttctcatgccaatactggtggaaaaatttcc 677

Qy        707 atttgttcgaaaatcagagagaacactggaatgttacgttcataacttactaaggattag 766
              |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db        678 atttgttcgaaaatcagagagaacactggaatgttacgttcatnacttactaaggattag 737

Qy        767 tgtatattttccaaccttgaggcatgaaattctggagcttattattgaaaaactactcaa 826
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        738 tgtatattttccaaccttgaggcatgaaattctggagcttattattgaaaaactactcaa 797

Qy        827 gtttggatgtgaatgcatcccggcagggtattgaaagatgctgaagaaacagcaactcaaac 886
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        798 gtttggatgtgaatgcatcccggcagggtattgaaagatgctgaagaaacagcaactcaaac 857

Qy        887 ttgtggtgggacagattccacggaaggattgtttaatatggatgaagatgaagaaactga 946
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        858 ttgtggtgggacagattccacggaaggattgtttaatatggatgaagatgaagaaactga 917

Qy        947 acatgaaacaaaggctggtcctgaacggctcgaccagatggtgcatcctgtagccgagcg 1006
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 acatgaaacaaaggctggtcctgaacggctcgaccagatggtgcatcctgtagccgagcg 977

Qy       1007 cctggacatcctgatgtctttggttttgtcctacatgaaggatgtctgctatgtagatgg 1066
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 cctggacatcctgatgtctttggttttgtcctacatgaaggatgtctgctatgtagatgg 1037

Qy       1067 taaggttgataacggcaaaacaaaggatctatatcgcgacctgataaacatctttgacaa 1126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 taaggttgataacggcaaaacaaaggatctatatcgcgacctgataaacatctttgacaa 1097

Qy       1127 actcctgttgccccacccatgcctcctgccatgtacagttttcatgtttacctctgtag 1186
              :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 actcctgttgccccacccatgcctcctgccatgtacagttttcatgtttacctctgtag 1157

Qy       1187 tttcaaattgggattcgcagaggcattttggaacatctctggaaaaaattgcaggaccc 1246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 tttcaaattgggattcgcagaggcattttggaacatctctggaaaaaattgcaggaccc 1217

Qy       1247 aagtaatcctgccatcatcaggcaggctgctggaaattatattggaagcttttggcaag 1306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1218 aagtaatcctgccatcatcaggcaggctgctggaaattatattggaagcttttggcaag 1277

Qy       1307 agctaaatttattcctcttattactgtaaaatcatgcctagatcttttggttaactggct 1366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1278 agctaaatttattcctcttattactgtaaaatcatgcctagatcttttggttaactggct 1337

Qy       1367 gcacatataccttaataaccaggattcggggaacaaaggcattctgcgatgttgctctcca 1426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 gcacatataccttaataaccaggattcggggaacaaaggcattctgcgatgttgctctcca 1397

Qy       1427 tggaccatttactcagcctgccaagctgtgttctacaccttttgttttagacacaagca 1486
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1398 tggaccatttactcagcctgccaagctgtgttctacncctttgttttagacacaagca 1457

Qy       1487 gcttttgagcggaaacctgaaagaaggtttgcagtatcttcagagtctgaatttttgagcg 1546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1458 gcttttgagcggaaacctgaaagaaggtttgcagtatcttcagagtctgaatttttgagcg 1517

Qy       1547 gatagtgatgagccagctaaatcccctgaagatttgcctgccctcagtggttaactttt 1606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1518 gatagtgatgagccagctaaatcccctgaagatttgcctgccctcagtggttaactttt 1577
```

```
Qy   1607 tgctgcaatcacaaataagtaccagctcgtcttctgctacaccatcattgaagaggaacaa 1666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1578 tgctgcaatcacaaataagtaccagctcgtcttctgctacaccatcattgaagaggaacaa 1637

Qy   1667 tcgccagatgctgccagtcattaggagtaccgctggaggagactcagtgcagatctgcac 1726
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1638 tcgccagatgctgccagtcattaggagtaccgctggaggagactcagtgcagatctgcac 1697

Qy   1727 aaacccgctggacacctttcttcccctttgatccctgtgtgctgaagaggtcaaagaaatt 1786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1698 aaacccgctggacacctttcttcccctttgatccctgtgtgctgaagaggtcaaagaaatt 1757

Qy   1787 cattgatcctatttatcaggtatgaggaagacatgagtgctgaagagctacaggagttcaa 1846
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1758 cattgatcctatttatcaggtgtgggaagacatgagtgctgaagagctacaggagttcaa 1817

Qy   1847 gaaacccatgaaaaaggacataagtggaaagatgaaggatgatgactttctgaaaggcgaagt 1906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1818 gaaacccatgaaaaaggacataagtggaaagatgaaggatgatgactttctgaaaggcgaagt 1877

Qy   1907 gcccagaatgataccgtgattgggatcacaccaagctcctttgacacgcatttccgaaag 1966
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1878 gcccagaatgataccgtgattgggatcacaccaagctcctttgacacgcatttccgaaag 1937

Qy   1967 tcttcaagtagtgtgggctcccacccgtgttgtacatgcaacccagtcccctctgacg 2026
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1938 tcttcaagtagtgtgggctcccacccgtgttgtacatgcaacccagtcccctctgacg 1997

Qy   2027 gcagaaatttgtgactgagatgtgacatttgggattccccat 2068
          |||||||||||||||||||||||||||||||||||||||||
Db   1998 gcagaaatttgtgactgagatgtgacatttgggattccccat 2039
```

RESULT 2
AAC76937
ID AAC76937 standard; cDNA; 1418 BP.
XX
AC AAC76937;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2492 polynucleotide sequence SEQ ID NO:4983.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR P-PSDB; AAB42728.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 4157-4158; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 1418 BP; 422 A; 275 C; 318 G; 403 T; 0 other;

Query Match 59.4%; Score 1228.4; DB 21; Length 1418;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 1249; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

```
Qy    670 cgtggtttctcatgccaatactggtgaaaaaatttccatttgttcgaaaatcagagagaa 729
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 cgtggtttctcatgccaatactggtgaaaaaatttccatttgttcgaaaatcagagagaa 60

Qy    730 cactgaaatgttacgttcataacttactaaggattagtgtatatttttccaa cctgaag 788
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     61 cactgaaatgttacgttcataacttactaaggattagtgtatatttttccaaccctgaag 120

Qy    789 catgaaattctgagagcttattattgaaaaactactcaagttggatgtgaatgcatcccag 848
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 catgaaattctgagagcttattattgaaaaactactcaagttggatgtgaatgcatcccag 180

Qy    849 cagggtattgaagatgctgaagaaacagcaactcaaacttgtggtgggacaaattccacg 908
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    181 cagggtattgaagatgctgaagaaacagcaactcaaacttttggtgggacaaattccacg 240

Qy    909 gaaggattgtttaatatggatgaagatgaagaaactgaacatgaaacaaagggctggtcct 968
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 gaaggattgtttaatatggatgaagatgaagaaactgaacatgaaacaaagggctggtcct 300

Qy    969 gaacggctcgaccagatggtgcatcctgtagccgaggcgcctggacatcctgatatctttg 1028
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 gaacggctcgaccagatggtgcatcctgtagccgaggcgcctggacatcctgatatctttg 360

Qy   1029 gttttgtcctacatgaaggatgtctgctatgtagatggtaaggttgataacggcgaaaaca 1088
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gttttgtcctacatgaaggatgtctgctatgtagatggtaaggttgataacggcgaaaaca 420

Qy   1089 aaggatctatatcgcgacctgataaacatctttgacaaactcctgttgcccaccatgcc 1148
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 aaggatctatatcgcgacctgataaacatctttgacaaactcctgttgcccaccatgcc 480

Qy   1149 tcctgccatgtacagtttttcatgtttacctctgtagtttcaaattggaattcgcagag 1208
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 tcctgccatgtacagtttttcatgttttacctctgtagtttcaaattgggattcgcagag 540

Qy      1209 gcatttttggaacatctctggaaaaaattgcaggacccaagtaatcctgccatcatcagg 1268
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 gcatttttggaacatctctggaaaaaattgcaggacccaagtaatcctgccatcatcagg 600

Qy      1269 caggctgctggaaattatattggaagcttttggcaagagctaaatttattcctcttatt 1328
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 caggctgctggaaattatattggaagcttttggcaagagctaaatttattcctcttatt 660

Qy      1329 actgtaaaatcatgcctagatctttggttaactggctgcacatataccttaataaccag 1388
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 actgtaaaatcatgcctagatctttggttaactggctgcacatataccttaataaccag 720

Qy      1389 gattcgggaacaaaggcattctgcgatgttgctctccatggaccatttactcagcctgc 1448
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 gattcgggaacaaaggcattctgcgatgttgctctccatggaccatttactcagcctgc 780

Qy      1449 caagctgtgttctacaccttgttttagacacaagcagctttgagcggaaacctgaaa 1508
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 caagctgtgttctacaccttgttttagacacaagcagctttgagcggaaacctgaaa 840

Qy      1509 gaaggtttgcagtatcttcagagtctgaatttgagcggatagtgatgagccagctaaat 1568
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 gaaggtttgcagtatcttcagagtctgaatttgagcggatagtgatgagccagctaaat 900

Qy      1569 cccctgaagatttgcctgccctcagtggttaactttttgctgcaatcacaaataagtac 1628
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 cccctgaagatttgcctgccctcagtggttaactttttgctgcaatcacaaataagtac 960

Qy      1629 cagctcgtcttctgctacaccatcattgagaggaacaatcgccagatgctgccagtcatt 1688
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 cagctcgtcttctgctacaccatcattgagaggaacaatcgccagatgctgccagtcatt 1020

Qy      1689 aggagtaccgctggaggagactcagtgcagatctgcacaaacccgctggacaccttcttc 1748
             ||||||||||||||||||||||||||||||| |||||||||||| |||||||||||||||
Db      1021 aggagtaccgctggaggagactcagtgcagacctgcacaaacccactggacaccttcttc 1080

Qy      1749 ccctttgatccctgtgtgctgaagaggtcaaagaaattcattgatcctatttatcaggta 1808
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db      1081 ccctttgatccctgtgtgctgaagaggtcaaagaaattcattgatcctatttatcaggtg 1140

Qy      1809 tgggaagacatgagtgctgaagagctacaggagttcaagaaacccatgaaaaaggacata 1868
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 tgggaagacatgagtgctgaagagctacaggagttcaagaaacccatgaaaaaggacata 1200

Qy      1869 gtggaagatgaagatgatgactttctgaaaggcgaagtgccccagaatgatacccgtgatt 1928
             ||||||||||||||||||||||||||||||||||| | ||||||||        || | |
Db      1201 gtggaagatgaagatgatgactttctgaaaggcgaaattccccagaaattagtagtaagt 1260

Qy      1929 gggatc 1934
             ||| ||
Db      1261 ggggtc 1266
```

RESULT 3
AAS44584
ID AAS44584 standard; DNA; 1498 BP.
XX
AC AAS44584;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human full-length polynucleotide sequence #9.
XX
KW Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW nervous system disorder; inflammatory disorder; cell differentiation; ds;
KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;
KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;
KW immunostimulant; analgesic; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200164834-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04926.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
PR 17-JUN-2000; 2000US-0597707.
PR 14-JUL-2000; 2000US-0616807.
PR 19-SEP-2000; 2000US-0664641.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
PI Drmanac R;
XX
DR WPI; 2001-589862/66.
DR P-PSDB; AAU27684.
XX
PT Novel polypeptides and nucleic acids obtained from cDNA libraries
PT prepared from various human tissues, for diagnosis, treatment of
PT cancer, neurological, inflammatory disorders and for use in arrays for
PT detection
XX
PS Claim 1; SEQ ID No 9; 153pp; English.
XX
CC Sequences AAS44576-AAS44919 represent full-length polynucleotides and
CC coding polynucleotides encoding polypeptides of the invention. The DNA
CC and protein sequences are useful for the treatment, diagnosis and
CC prevention of various types of disorder in a mammalian subject such as a
CC human, dog, monkey, mouse, hamster or rat. The disorders include cancers
CC such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such
CC as multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC bowel disease. The sequences exhibit activity relating to angiogenesis,
CC cell proliferation, cell differentiation, stem cell growth factor,
CC activin or inhibin. Therefore, they can be used to manipulate stem cells
CC in culture to give rise to neuroepithelial cells that can be used to
CC augment or replace cells damaged by illness, accidental damage or genetic
CC disorders. The sequences may also be used for regeneration of bone,
CC cartilage, tendons and ligaments and in tissue repair and burn healing.
CC Note: Some sequences for this patent did not form part of the printed
CC specification, but were obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1498 BP; 417 A; 332 C; 325 G; 424 T; 0 other;

```
  Query Match             48.9%;  Score 1010.4;  DB 22;  Length 1498;
  Best Local Similarity   98.8%;  Pred. No. 4.3e-287;
  Matches 1028;  Conservative    0;  Mismatches   11;  Indels    1;  Gaps    1;

Qy       735 gaatgttacgttcataacttactaaggattagtgtatattttccaaccttgaggcatgaa 794
             :|| |||||||||||||||||  ||||||||||||||||||||||||| ||||||||||
Db         4 gaaagttacgttcataacttacagaggattagtgtatattttccaacc-tgaggcatgaa 62

Qy       795 attctggagcttattattgaaaaactactcaagttggatgtgaatgcatcccggcagggt 854
             :|||||  ||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        63 gttctgaggcttattatggaaaaactactcaagttggatgtgaatgcatcccggcagggt 122

Qy       855 gttgaagatgctgaagaaacagcaactcaaacttgtggtgggacagattccacggaagga 914
```

```
Db    123 attgaagatgctgaagaaacagcaactcaaacttgtggtgggacagattccacggaagga 182

Qy    915 ttgtttaatatggatgaaagatgaagaaactgaacatgaaacaaaggctggtcctgaacgg 974

Db    183 ttgtttaatatggatgaagatgaagaaactgaacatgaaacaaaggctggtcctgaacgg 242

Qy    975 ctcgaccagatggtgcatcctgtagccgagcgcctggacatcctgatgtctttggttttg 1034

Db    243 ctcgaccagatggtgcatcctgtagccgagcgcctggacatcctgatgtctttggttttg 302

Qy   1035 tcctacatgaaggatgtctgctatgtagatggtaaggttgataacggcaaaacaaaggat 1094

Db    303 tcctacatgaaggatgtctgctatgtagatggtaaggttgataacggcaaaacaaaggat 362

Qy   1095 ctatatcgcgacctgataaacatctttgacaaactcctgttgcccacccatgcctcctgc 1154

Db    363 ctatatcgcgacctgataaacatctttgacaaactcctgttgcccacccatgcctcctgc 422

Qy   1155 catgtacagttttttcatgtttttaccttctgtagtttcaaattgggattcgcagaggcattt 1214

Db    423 catgtacagttttttcatgtttttaccttctgtagtttcaaattgggattcgcagaggcattt 482

Qy   1215 ttggaacatctctgaaaaaaattgcaggacccaagtaatcctgccatcatcaggcaggct 1274

Db    483 ttggaacatctctggaaaaaaattgcaggacccaagtaatcctgccatcatcaggcaggct 542

Qy   1275 gctggaaattatattgaaagcttttttggcaaagagctaaatttattccttcttattactgta 1334

Db    543 gctggaaattatattggaagcttttttggcaaagagctaaatttattccttcttattactgta 602

Qy   1335 aaatcatgcctagatcttttggttaactggctgcacatatacctttaataaccagaattcg 1394

Db    603 aaatcatgcctagatcttttggttaactggctgcacatatacctttaataaccagaattcg 662

Qy   1395 ggaacaaaggcattctgcgatgttgctctccatggaccatttttactcagcctgccaagct 1454

Db    663 ggaacaaaggcattctgcgatgttgctctccatggaccatttttactcagcctgccaagct 722

Qy   1455 gtgttctacacctttgtttttagacacaagcagcttttgagcggaaacctgaaagaaggt 1514

Db    723 gtgttctacacctttgtttttagacacaagcagcttttgagcggaaacctgaaagaaggt 782

Qy   1515 ttgcagtatcttcagagtctgaattttgaagcggatagtgatgaagccagctaaatccccctg 1574

Db    783 ttgcagtatcttcagagtctgaattttgaagcggatagtgatgaagccagctaaatccccctg 842

Qy   1575 aagatttgcctgccctcagtggttaacttttttgctgcaatcacaaataagtaccagctc 1634

Db    843 aagatttgcctgccctcagtggttaacttttttgctgcaatcacaaataagtaccagctc 902

Qy   1635 gtcttctgctacaccatcattgaagaggaacaatcgccagatgctgccagtcattaggagt 1694

Db    903 gtcttctgctacaccatcatcgagaggaacaatcgccagatgctgccagtcattaggagt 962

Qy   1695 accgctggaggagactcagtgcagatctgcacaaacccgctggacaccttcttccccttt 1754

Db    963 accactggaggagactcagtgcagacctgcacaaacccactggacaccttcgtccccttt 1022

Qy   1755 gatccctgtgtgctgaagag 1774

Db   1023 gatccctgtgtgctgaagag 1042


RESULT   4
AAS01562
ID   AAS01562 standard; cDNA; 1423 BP.
XX
AC   AAS01562;
XX
DT   18-JUL-2001  (first entry)
XX
DE   Human secretory molecule cDNA sptm #52.
XX
KW   Human; secretory molecule; sptm; SPTM; library screening; gene therapy;
KW   cell signalling; cell proliferative disorder; atherosclerosis; cancer;
KW   immune system disorder; AIDS; neurological disorder; Alzheimer's disease;
KW   nervous system disease; mental retardation; developmental disorder;
KW   neuromuscular disorder; microarray; Incyte ID number 4287452dec; ss.
XX
OS   Homo sapiens.
XX
PN   WO200123558-A2.
XX
PD   05-APR-2001.
XX
PF   19-SEP-2000; 2000WO-US25610.
XX
PR   28-SEP-1999;   99US-0156624.
PR   28-SEP-1999;   99US-0156625.
PR   02-DEC-1999;   99US-0168611.
PR   02-DEC-1999;   99US-0168613.
PR   02-DEC-1999;   99US-0168614.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Hodgson DM,  Lincoln SE,  Rosso FD,  Spiro PA,  Banville SC;
PI   Bratcher SR,  Dufour GE,  Cohen HJ,  Rosen BH,  Shah P,  Chalup MS;
PI   Hillman JL,  Jones AL,  Yu JY,  Greenawalt LB,  Panzer SR;
PI   Roseberry AM,  Wright RJ,  Chen W,  Liu TF,  Yap PE,  Stockdreher TK;
PI   Amshey S,  Fong WT;
XX
DR   WPI; 2001-258134/26.
XX
PT   New secretory polynucleotides (sptm) and the polypeptides they encode,
PT   for use in inducing antibodies and screening libraries of compounds -
XX
PS   Claim 1; Page 152; 161pp; English.
XX
CC   The present sequence for human secretory molecule cDNA sptm #52
CC   (Incyte ID number 4287452dec) is 1 of 63 novel sptm cDNA sequences
CC   (AAS01511-AAS01573) which encode for the secretory polypeptides SPTM. The
CC   sptm polynucleotides are useful for screening a compound for
CC   effectiveness in altering expression of a target polynucleotide, where
CC   the target polynucleotide comprises sptm. Sptm is also useful in a method
CC   for assessing the toxicity of a test compound. Sptm and its fragments or
CC   complementary sequences, may be used to identify the presence of and/or
CC   determine the degree of similarity between two nucleic acid sequences.
CC   Sptm can also be used for a variety of diagnostic and therapeutic
CC   purposes, e.g. diagnosing a particular condition, disease or disorder
CC   associated with cell signalling, such conditions include cell
CC   proliferative disorders such as atherosclerosis, and cancers including
CC   leukaemia, an immune system disorder e.g. acquired immunodeficiency
CC   syndrome (AIDS), a neurological disorder such as epilepsy or Alzheimer's
CC   disease, nutritional and metabolic disease of the nervous system, mental
CC   retardation and other developmental disorders, and muscular dystrophy and
CC   other neuromuscular disorders. Sptm can also be used to design probes
CC   useful in diagnostic assays, which may be used to monitor the progress of
CC   conditions or disorders associated with abnormal levels of expression of
CC   sptm. In addition sptm encoding SPTM may be used for somatic or germline
CC   gene therapy, for inducing antibodies, or in microarrays.
XX
SQ   Sequence 1423 BP; 417 A; 247 C; 320 G; 439 T; 0 other;


  Query Match             44.2%;  Score 894;  DB 22;  Length 1423;
  Best Local Similarity   80.9%;  Pred. No. 8.7e-253;
  Matches 1191;  Conservative    0;  Mismatches   10;  Indels  271;  Gaps    2;

Qy    157 ggatttcaaatatgcgtgcattagaagaatgacttttttcaattctcccccaaaaaaactg 216

Db      1 ggatttcaaatatgcgtgcattagaagaatgatttttttcaattctcccccaaaaaaactg 251

Qy    217 ttcggtttggatggaactgtgacagaaagtcttgctgaaagtacaaaaaaggatgaaacaaatg 276

Db      2 ttcggtttggatggaactgtgacagaaagtcttgctgaaagtacaaaaaaggatgaaacaaatg 311
```

```
Qy    277 actttgagttgttgaagaaccagctgttagatccagacataaaggatgaccagatcatca 336
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    312 actttgagttgttgaagaaccagctgttagatccagacataaaggatgaccagatcatca 371

Qy    337 actgagctgctagaattccgttcttctatcatgtacttgacaaaagactttgaagcaactta 396
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    372 actgagctgctagaattccgttcttctatcatgtacttgacaaaagactttgaagcaactta 431

Qy    397 tcagtattatattaagattgcttgttgaataggaagtcaaacagtagtgtggaagagtatt 456
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    432 tcagtattatattaagattgcttgttgaataggaagtcaaacagtagtgtggaagagtatt 491

Qy    457 tggcttttcttggtaatcttgtatcagcacagactgttttccttcagaccgtgtctcagca 516
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    492 tggctttcttggtaatcttgtatcagcacagactgttttccttcagaccgtgtctcagca 551

Qy    517 tgattgcttcccatttttgtgcctccccgaagtgatcattaaggaaggcgatgtagatgttt 576
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    552 tgattgcttcccatttttgtgcctccccgaagtgatcattaaggaaggcgatgtagatgttt 611

Qy    577 cagattctgatgatgaaagatgataatcttcctgcaaatttgacacatgtcacagaggcct 636
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    612 cagattctgatgatgaaagatgataatcttcctgcaaatttgacacatgtcacagaggcct 671

Qy    637 tgcaaataatagcaagatatgtaccatcgacacccgtggtttctcatgccaatactggtgg 696
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    672 tgcaaataatagcaagatatgtaccatcgacacccgtggtttctcatgccaatactggtgg 731

Qy    697 aaaaatttccatttgttcagaaaatcagagagaacactggaatgttacgttcataacttac 756
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    732 aaaaatttccatttgttcagaaaatcagagagaacactggaatgttacgttcataacttac 791

Qy    757 taaggattagtgtatatttccaacttgaggcatgaaattctggagcttattattgaaa 816
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    792 taaggattagtgtatatttccaacttgaggcatgaaattctggagcttattattgaaa 851

Qy    817 aactactcaagttggatgtgaatgcatcccggcaggtattgaagatgctgaagaaacag 876
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    852 aactactcaagttggatgtgaatgcatcccggcaggtattgaagatgctgaagaaacag 911

Qy    877 caactcaaacttgtggtggaacagattccacggaaggattgtttaatatgaatgaagatg 936
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    912 caattcaaacttgtggtggaacagattccacggaaggattgtttaata---------- 959

Qy    937 aagaaactgaacatgaaacaaaggctggtcctgaacggctcgaccagatggtgcatcctg 996

Db    960 ----- ---------- ---------- ---------- 959

Qy    997 tagccgagcgcctggacatcctgatgtctttggtttgtcctacatgaaggatgtctgct 1056

Db    960 --- ---------- ---------- --- 959

Qy   1057 atgtagatggtaaggttgataacggcgaaaacaaaggatctatatcgcgacctgataaaca 1116

Db    960 ---------- ---------- ---------- 959

Qy   1117 tctttgacaaactcctgttgcccacccatgccctcctgccatgtacagttttcatgtttt 1176

Db    960 - ------ ---------- ---------- 959

Qy   1177 acctctgtagtttcaaattggggattcgcagagg-catttttggaacatctctggaaaaaa 1235
                    ||||||||||||||| |||||||||||||||| ||||||||
Db    960 ---- ---- tgggattcgcagaggccatttttggaacatctttggaaaaaac 1001

Qy   1236 ttgcagggacccaagtaatcctgccatcatcagccagggctgctggaaattatattgggaagc 1295
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1002 ttgcagggatccaagtaatcctgccatcatcagccagggctgctggaaattatattgggaagc 1061

Qy   1296 ttttgccagcagagctaaatttattcctcttattactgtaaaatcatgccctagatcttttg 1355
          ||||||||||||||||||||||| |||||||||||||||| ||||||||||||||||||
Db   1062 ttttggccagcagagctaaatttatttctcttattactgtaaaatcatgccctagatcttttg 1121

Qy   1356 gttaactgactgcagcatatacctttaataaccagcattcgggaaacaaaagccattctgcgat 1415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1122 gttaactgactgcagcatatacctttaataaccagcattcgggaaacaaaagccattctgcgat 1181

Qy   1416 gttgctctccatggaccattttactcagcctgccaagctgtgttctacacctttgttttt 1475
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1182 gttgctctccatggaccattttactcagcctgccaagctgtgttctacacctttgttttt 1241

Qy   1476 agacacaagcagctttgagcggaaacctgaaaagaaagtttgcagtatcttcaaaatctg 1535
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1242 agacacaagcagctttgagcggaaacctgaaaagaaagtttgcagtatcttcaaagtctg 1301

Qy   1536 aatttgaagcggatagtgatgaagccagctaaatccgctgaagatttgcctgccctcagtg 1595
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1302 aatttgaagcggatagtgatgaagccagctaaatccgctgaagatttgcctgccctcagtg 1361

Qy   1596 gttaacttttttgctgcaatcacaaataaagta 1627
          |||||||||||||||||||||||||||||| | | |
Db   1362 gttaacttttttgctgcaatcacaaagatgaa 1393


RESULT 5
AAS92255
ID   AAS92255 standard; cDNA; 2410 BP.
XX
AC   AAS92255;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #28059.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG28068.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 1; SEQ ID NO 28059; 104pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
```

```
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2410 BP; 672 A; 437 C; 547 G; 754 T; 0 other;


  Query Match             30.9%;  Score 639.4;  DB 23;  Length 2410;
  Best Local Similarity   96.8%;  Pred. No. 1.4e-177;
  Matches  706;  Conservative    0;  Mismatches   16;  Indels    7;  Gaps    5;

Qy    1347 gatcttttgqttaactggctgcacatatacottaataaccaggattcqggaacaaaggca 1406
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 gatcttttggttaactggctgcacatatacottaataaccaggattcgggaacaaaggca 60

Qy    1407 ttctgcgatgttgctctccatggaccattttactcagcctgccaagctgtgttctacacc 1466
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ttctgcgatgttgctctccatggaccattttactcagcctgccaagctgtgttctacacc 120

Qy    1467 tttgtttttagacacaagcagcttttgagcggaaacctgaaagaaggtttgcagtatctt 1526
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 tttgtttttagacacaagcagcttttgagcggaaacctgaaagaaggtttgcagtatctt 180

Qy    1527 cagaagtctgaattttgaagcggataqtgatgaagccagctaaatcccctgaagat-ttgcct 1585
           ||||||||||||||||| |||||||||||||||||||||||||||||||||| | || |
Db     181 cagaagtctgaattttgaagtggatagtgatgaggccagctaaatcccctgaagatatggctt 240

Qy    1586 gccctcagtggttaacttttttgctgcaatcacaaataagtaccagctcgtcttctgcta 1645
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 gccctcagtggttaacttttttgctgcaatcacaaataagtaccagctcgtcttctgcta 300

Qy    1646 caccatcattgagaggaacaatcgccagatgctgccagtcattaggagtaccgctggagg 1705
           ||||||||| |||||||||||||||||||||||||||||||||||||||||| |||||||
Db     301 caccatcatcgagaggaacaatcgccagatgctgccagtcattaggagtaccactggagg 360

Qy    1706 agactcagtgcagatctgcacaaacccgctggacaccttcttcccctttgatccctgtg 1764
           |||||||||||||| |||||||||| |||||||||||||||||||| ||||||||||
Db     361 agactcagtgcagacctgcacaaacccactggacaccttcttcccctttggatccctgtg 420

Qy    1765 tgctgaagaggtcaaagaaatt-cattgatcctatttt---atcaggtatgqggaagacatg 1820
           |||||||||||||||||||| |||||||||||||     |    ||||||||||
Db     421 tgctgaagaggtcaaagaaattccattgatcctattttatccaggtatgqggaagacatg 480

Qy    1821 agtgctgaaagagctacaggagttcaagaaacccatgaaaaaggacatagtggaagatgaa 1880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 agtgctgaagagctacaggagttcaagaaacccatgaaaaaggacatagtggaagatgaa 540

Qy    1881 gatgatgactttctgaaaggcgaagtgccccagaatgataccgtgattgggatcacacca 1940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 gatgatgactttctgaaaggcgaagtgccccagaatgataccgtgattgggatcacacca 600

Qy    1941 agctcctttgacacgcatttccgaagtccttcaagtagtgtgggctccccacccgtgttg 2000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 agctcctttgacacgcatttccgaagtccttcaagtagtgtgggctccccacccgtgttg 660

Qy    2001 taca-tgcaacccagtcccctctgacggcagaaatttgtgactgagatgtgacatttggg 2059
           |||| ||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     661 tacactgcaacccagtcccctctgacggcagaaatttgcgactgagatgtgacatttggg 720

Qy    2060 attcccat 2068
           ||||||||
Db     721 attcccat 729


RESULT   6
AAS92253
ID   AAS92253 standard; cDNA; 776 BP.
XX
AC   AAS92253;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #28057.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG28056.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 1; SEQ ID No 28057; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 776 BP; 206 A; 157 C; 198 G; 215 T; 0 other;


  Query Match             27.5%;  Score 568.6;  DB 23;  Length 776;
  Best Local Similarity   97.8%;  Pred. No. 5.6e-157;
  Matches  629;  Conservative    0;  Mismatches    9;  Indels    5;  Gaps    5;

Qy       4 gaggctgtggctggaaggaggctgggcatccggcctgaggcgcagcggtcgcgttagttcg 63
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133 gaggctgtggctggaaggaggctgggcatccggcctgaggcgcagcggtcgcgttagttcg 192

Qy      64 gccgaatgggcggcaccgctgcttcacacg-catttgccgggaggatgcgggcggttcgtcc 122
           |||||||||||||||||||||||||||||  ||||| ||||||||||||||||||||||
Db     193 gccgaatgggcggcaccgctgcttcacacgttgtttgtcgggaggatgcgggcggttcgtcc 252

Qy     123 tctgcagttaaggaagctgggcgcgtcg-aggactgggatttcaaatatgcgtgcattagg 181
           ||||||| |||| |||| |||||||| ||||||||||||||||||||||||||||||||
Db     253 tctgcagtcaaggcgctgttcgcgtcgtaggactgggatttcaaatatgcgtgcattagg 312
```

```
Qy     182 gaatgacttttttcaattctccccccaaaaaaactgttcggtttgqtgqaactgtgacaga 241
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     414 gaatgattttttcaattctccccccaaaaaaactgttcggtttgqtgqaactgtgacaga 372

Qy     242 agtcttgctgaaqtacaaaaaaggqtgaaacaaatgacttgaagttgttgaaagaaccaget 301
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     373 agtcttgctgaaqtacaaaaaaggqtgaaacaaatgacttgaagttgttgaaagaaccaget 432

Qy     302 gttagatccagacataaaggatga-ccagatcatcaact qgctgctagaattccgttct 359
           |||||||||||||||||||||| |||||||||||||| ||||||||||||||||||||
Db     433 gttagatccagacataaaggatgaccccagatcatcaactgggctgctagaattccgttct 492

Qy     360 tctatcatgtacttgaccaaaaagacttgaagcaacttatcagtattatatttaagattgcc- 418
           ||| |||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     493 tctgtcatgtacttgaccaaaaagacttgaagcaacttatcagtattatatttaagattgcct 552

Qy     419 ttgattgaataggtcaaacagtagtgaaagagtatttgggtttttcttgqtaatcttgt 478
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     553 ttgattgaataggtcaaacagtagtgaaagagtatttgggtttttcttgqtaatcttgt 612

Qy     479 atcagcacagactgtttcctcagaccgtgtctcagcatgattgcttcccatttgtgcc 538
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     613 atcagcacagactgtttcctcagaccgtgtctcagcatgattgcttcccatttgtgcc 672

Qy     539 tccccgaqtgatcattaaggaaaggcgatgtagatgtttcagattctgatgatgaaagatga 598
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     673 tccccgaqtgatcattaaggaaaggcgatgtagatgtttcagattctgatgatgaaagatga 732

Qy     599 taatcttcctgcaaatttgacacatgtcacagaqccttgcaa 641
           |||||||||||||||||||||||||||||||||||||||||
Db     733 taatcttcctgcaaatttgacacatgtcacagaqccttgcaa 775


RESULT  7
AAS87643
ID   AAS87643 standard; cDNA; 1260 BP.
XX
AC   AAS87643;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #23447.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE ) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; AB023456.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 1; SEQ ID No 23447; 104pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1260 BP; 344 A; 227 C; 269 G; 420 T; 0 other;


  Query Match             26.6%;  Score 550.8;  DB 23;  Length 1260;
  Best Local Similarity   89.0%;  Pred. No. 1.3e-151;
  Matches  696;  Conservative    0;  Mismatches   77;  Indels    9;  Gaps    9;

Qy       4 gaagctgtgagctgaaaagaagctgggcatccagccctgaggcgcagcggtcgcgttaagttcg 63
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     132 gaagctgtgagctgaaaagaagctgggcatccagccctgaggcgcagcggtcgcgttaagttcg 192

Qy      64 gcccaatgagcggcaccgcgctgcttcacacg-cgtttgcccggagagatgcgggccgcttcgtcc 122
           |||||||||||||||||||||||||||||  ||||| ||||||||||||||||||||||||
Db     193 gcccaatgagcggcaccgcgctgcttcacacgttgtttgtccggagagatgcgggccgcttcgtcc 252

Qy     123 tctgcagttaaagaagctgggcgcgtcg-aggactgaggatttcaaatatgcgtgcattaga 181
           ||||||| |||| ||||  ||||||| ||||||||||||||||||||||||||||||||||
Db     253 tctgcagttcaagaagctgggcgctgtcgtagggactgaggatttcaaatatgcgtgcattaga 312

Qy     182 gaatgacttttttcaattctccccccaaaaaaaactgttcggtttgqtgqaactgtgacaga 241
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     313 gaatgatttttttcaattctccccccaaaaaaaactgttcggtttgqtgqaactgtgacaga 372

Qy     242 agtcttgctgaaqtacaaaaaaggqtgaaacaaatgacttttgaagttgttgaaagaaccaget 301
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     373 agtcttgctgaaqtacaaaaaaggqtgaaacaaatgacttttgaagttgttgaaagaaccaget 432

Qy     302 gttagatccagacataaaggatga ccagatcatcaact-gqctgctagaattccgttct 359
           |||||||||||||||||||||||| |||||||||||||| |||||||||||||||||||
Db     433 gttagatccagacataaaggatgacccagatcatcaactgggctgctagaattccgttct 492

Qy     360 tctatcatgtacttga-caaaaagacttgaag caacttatcagtattatatttaagattgc 417
           ||| |||||||||||| |||||||||||||| ||||||||||||||||||| |||||||
Db     493 tctgtcatgtacttgaccaaaaagacttgaagccaacttatcagtattatatttaagattgc 552

Qy     418 c-ttgattgaataggtcaaacagtagtgaaagagtatttgggtttttcttggtaatct 475
           | ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     553 ctttgattgaataggtcaaaacagtagtgaaagagtatttgggtttttcttgqtaatct 612

Qy     476 tgtatcagcacagactgtttcctcagaccgtgtctcagcatgattgctt ccgattttg 534
           |||||||||| |||||||||||||||||||||||||||||||||||| |||||||||
Db     613 tgtatcagcatagactgtttcctcagaccgtgtctcagcatgattgcttcccgattttg 672

Qy     535 tgccttcccgaagtgatcattaaggaaaggcgatgtagatgtttcagattctgatgatgaaag 594
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     673 tgccttcccgaagtgatcattaaggaaaggcgatgtagatgtttcagattctgatgatgaaag 732

Qy     595 atgataatcttcctgcaaatttgacacatgtcacagaqccttgcaaataataqcaaaat 654
           |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db     733 atgataatcttcctgcaaatttgacacatatcacagaqccttgcaaataataqcaaaat 792

Qy     655 atgtaccatcgacaccgtcgtttctcatgccaatactgctgqaaaaatttccatttattc 714
```

```
Db    793 atgtaccatcatgagtatactttttccttattttgaatgtttaattctcaaqaaaattgta 852

Qy    715 qaaaatcagagagaacactqgaatgttacgttcataacttactaaggattaqtqtatatt 774
Db    853 atcaattagtaaaaattataaaatgttaatagtattaaagcttgagtcttacattgcatt 912

Qy    775 tt 776
Db    913 tt 914


RESULT   8
AAS80955/c
ID   AAS80955 standard; cDNA; 655 BP.
XX
AC   AAS80955;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #16759.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG16768.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 1; SEQ ID No 16759; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 655 BP; 207 A; 134 C; 115 G; 199 T; 0 other;
```

```
  Query Match             23.1%;  Score 477.8;  DB 23;  Length 655;
  Best Local Similarity   90.2%;  Pred. No. 3e-130;
  Matches  578;  Conservative    0;  Mismatches   57;  Indels    6;  Gaps    6;

Qy    108 qcqqccqcttcqtcctctqcaqttaaqaaqctqqqcqccqtcq-aqqactqqqatttcaaa 166
Db    655 GCGGCCGCTTCGTCCTGTGCAGTCAAGACGCTGTTCGCGTCGTAGGACTGGGATTTCAAA 596

Qy    167 tatqcqtqcattaqaqaatqactttttcaattctccccccaaqaaaaaactqttcqqtttqq 226
Db    595 TATGCGTGCATTAGAGAATGATTTTTTCAATTCTCCCCCAAGAAAAACTGTTCGGTTTGG 536

Qy    227 tqqaactqtqacaqaaqtcttqctqaaqtacaaaaaqqqtqaaacaaatqactttqaqtt 286
Db    535 TGGAACTGTGACAGAAGTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTT 476

Qy    287 qttqaaqaaccaqctqttaqatccaqacataaaqqatqa-ccaqatcatcaactqqctqc 345
Db    475 GTTGAAGAACCAGCTGTTAGATCCAGACATAAAGGATGACCCAGATCATCAACTGGCTGC 416

Qy    346 tagaattccattcttctatcatqtacttqa-caaaaqactttqaq-caacttatcaatat 403
Db    415 TAGAATTCCGTTCTTCTGTCATGTACTTGACCAAAAGACTTTGAGCCAACTTATCAGTAT 356

Qy    404 tatattaaqattqccttqqttqaaataqaaqtc-aaacaqtaqtqqaaqaqtatttqqctt 462
Db    355 TATATTGAGATTGCCTTGGTTGAATAGAAGTCAAAACAGTAGTGGAAGAGTATTTGGCTT 296

Qy    463 ttcttqqtaatcttqtatcaqcacaqactqttttcctcaqaccqtqtctcaqcatqattq 522
Db    295 TTCTTGGTAATCTTGTATCAGCATAGACTGTTTTCCTCAGACCGTGTCTCAGCATGATTG 236

Qy    523 ctt-cccatttqtqcctcccqaqtqatcattaaqqaaqqcqatqtaqatqtttcaqat 581
Db    235 CTTCCCCATTTTGTGCCTCCCCGAGTGATCATTAAGGAAGGCGATGTAGATGTTTCAGAT 176

Qy    582 tctqatqatqaaqatqataatcttcctqcaaattttqacacatqtcacaqaqcctqcaa 641
Db    175 TCTGATGATGAAGATGATAATCTTCCTGCAAATTTTGACACATATCACAGAGCCTTGCAA 116

Qy    642 ataataqcaaqatatqtaccatcqacaccqtqqttttctcatqccaatactqqtqqaaaaa 701
Db    115 ATAATAGCAAGATATGTACCATCATGAGTATACTTTTCCTTATTTTGAATGTTTAATTCT 56

Qy    702 tttccatttgttcqaaaatcaqaqaqaacactqqaatqtta 742
Db     55 CAACAAAATTGTAATCAATTAGTAAAAATTATAAAATGTTA 15


RESULT   9
AAS87641
ID   AAS87641 standard; cDNA; 3169 BP.
XX
AC   AAS87641;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #23445.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
```

```
PR   23 AUG 2000; 2000US-0649167.
XX
PA   (HYSE ) HYSEQ INC.
XX
PI   Drmanac RT, Lin C, Tang YT;
XX
DR   WPI; 2001-649362/73.
DR   P-PSDB; AB624454.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 1; SEQ ID No 23445; 104pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4169 BP; 812 A; 841 C; 779 G; 737 T; 0 other;
```

```
  Query Match             22.5%;  Score 466.2;  DB 23;  Length 4169;
  Best Local Similarity   64.6%;  Pred. No. 2e-126;
  Matches  933;  Conservative    0;  Mismatches  213;  Indels  299;  Gaps    4;

Qy      69 atqqcqqcaccqctqcttcacacqcqtttqcccqqqaqatqcqqccqcttcqtcctctqca 128
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 atqqcqqcaccqctqcttcacacqcqtttqcccqqqaqatqcqqccqcttcqcccctctqca  60

Qy     129 qttaaqaaqctqqqcqcqtcqaqqactqqqatttcaaatatqcqtqcattaqaqaaatgac 188
           || |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 qtcaaqatqctqqqcqcqtcqaqqactqqqatttcaaatatqcqtqcattaqaqaaatgat 120

Qy     189 ttttcaattctccccccaaqaaaaactqttcqqtttqqtqaaactqtqacaqaaqtcttq 248
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     121 ttttcaattctccccccaaqaaaaactqttcaqtttqqtqaaactqtqacaqaaqtcttq 180

Qy     249 ctqaaqtacaaaaaqqqtqaaacaaatqacttqaqttqttqaaqaaccaqctqttagat 308
           ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ctqaaqtacaaaacqqqtqaaacaaatqacttqaqttqttqaaqaaccaqctqttagat 240

Qy     309 ccaqacataaaqqatqaccaqatcatcaactqqctqctaqaaattccqttcttctatcatq 368
           ||||||||||||  |       |  ||    | || |                  |||
Db     241 ccaqacataaaqqtqqtqtqcccccqccatqccqatqatcctqaqqqaqqqqqaqqqccatt 300

Qy     369 tacttqaccaaaqacttqaqcaactattcaqtattatattaaqattqccttqqttqaat 428
                 | |||| | |  |||    |  |:   ||  | |       |
Db     301 qtqqqcqaqtqtqccttcatcccctteqtqqcttcaqttqtqtcqtcaaaccaqaqtqqc 360

Qy     429 aqaaqtcaaacaqtaqtqqaaqaqtatttqqcttttcttqqtaatcttgtatcaqcacaq 488
           || |  |  || |   |  |                    | ||     |  | |
Db     361 ctcaqcccqaaqqttqctaaqctq  --- - --  -------  cctqcqqtqqqtcqq 399
```

```
Qy     489 actqtttcctcaqaccqtqtctcaqcatqattqcttcccatttttqtqcctccccqaqtq 548
           | ||   |  | | ||      ||||  |    | |||       |  | || || |
Db     400 cccqtqqcctqcccqaccccacatcaqccccqcctctqccacaaccccccaaccccaqcq 459

Qy     549 atcattaaqqaaqqqcqatqtaqatqtttcaqattctqatqatqaaqatqataatcttcct 608
             ||  ||    |       | | |
Db     460 caqcttctqqctcaacteatctqctttqca  -------- -                  488

Qy     609 qcaaatttqtcacatqtcacaqaqccttqcaaataataqccaaqatatqtaccatcqaca 668
                                                                  ||||
Db     489 ---------- --------  --------  -  -------qqaca 494

Qy     669 cqtqqtttctcatqccaatactqqtqqaaaaattttccatttqttcqaaaatcaqaqaqa 728
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     495 cqtqqtttctcatqccaatactqqtqqaaaaattttccatttqttcqaaaatcaqaqaqa 554

Qy     729 acactqaaatqttacqttcataacttactaaqqattaqtqtatatttttccaaccttqaaq 788
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     555 acactqaaatqttacqttcataacttactaaqqattaqtqtatatttttccaaccttqaaq 614

Qy     789 catqaaaattctqqaqcttattattqaaaaactactccaaqttqqatqtqaatqcatcccqq 848
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     615 catqaaaattctqqaqcttattattqaaaaactactccaaqttqqatqtqaatqcatcccqq 674

Qy     849 caqqqtattqaaaqatqctqaaqaaaacaqccaaatcaaaacttqtqtqtqqqacaqatttccacq 908
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     675 caqqqtattqaaaqatqctqaaqaaaacaqccaaatcaaaacttqtqtqtqqqacaqatttccacq 734

Qy     909 gaaqqattqtttaatatqqatqaaqatqaaqaaaactqaacatqaaacaaqqqctqqtcct 968
           ||||||||||| ||||||||
Db     735 gaaqqattqtqtaatatqq-------  -------- --  ---------- -------- 752

Qy     969 gaacqqctcqaccaqatqqtqcatcctqtaqccqaqcqcctqqacatcctqatqtctttq 1028
Db     753 ---------------------- ------------ -------  - --- --    752

Qy    1029 gtttqtcctacatqaaqqatqtctqctatqtaqatqqtaaqqttqataacqqcaaaaca 1088
           || |   | || || ||     |  ||| || |  | |   | | |  |  ||||
Db     753 --ttaqcaqtttattaatqaaaqtqqaqatqaaqtttatcataatcaaaqqqtqaaaaca 810

Qy    1089 aaqqatctatatcqcqacctqataaacatctttqacaaactcctqttqcccaccccatqcc 1148
Db     811 t- -      -------      --------  -----------   ----          811

Qy    1149 tcctqccatqtacaqttttcatqtttacctctqtaqttcaaattqaqattcqcaqaq 1208
                                          ||         |||  || ||||||||||||
Db     812 -          ----------          etaqtqctqctcatctttqqattcqcaqaq 841

Qy    1209 gcattttqaaacatctctqaaaaaaattqcaqqacccaaqtaatcctqccatcatcaqq 1268
           ||||||||||||| ||| |||||||||||||||| ||||||||||||||||||||||||
Db     842 acattttqaaacctctttqaaaaaaattqcaqqatccaaqtaatcctqccatcatcaqq 901

Qy    1269 caqqctqctqqaaattatattqqaaqctttttqqcaaqaqctaaattttatcctcttatt 1328
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     902 caqqctqctqqaaattatattqqaaqctttttqqcaaqaqctaaattttatcctcttatt 961

Qy    1329 actqtaaaatcatqcctaqatcttttqqttaactqqctqcacatatacctttaataaccaq 1388
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     962 actqtaaaatcatqcctaqatcttttqqttaactqqctqcacatatacctttaataaccaq 1021

Qy    1389 gatteqqqaacaaaqqcattctqcqatqttqctctccatqqaccatttttactcaqcctqc 1448
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1022 gatteqqqaacaaaqqcattctqcqatqttqctctccatqqaccatttttactcaqcctqc 1081

Qy    1449 caaqctqtqttctcacacctttqttttttaqacacaaqcaqcttttqaqcqqaaacctqaaa 1508
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1082 caaqctqtqttctcacacctttqttttttaqacacaaqcaqcttttqaqcqqaaacctqaaa 1141

Qy    1509 aaaq 1513
           |-|||
Db     1142 aaaq 1146
```

RESULT 10
ABA83038
ID ABA83038 standard; DNA; 1461 BP.
XX
AC ABA83038;
XX
DT 05-FEB-2002 (first entry)
XX
DE Human transcription factor TRFX-65 coding sequence.
XX
KW Human; transcription factor; TRFX; cell proliferative disease;
KW autoimmune disease; inflammation; neurological disease;
KW developmental disorder; cancer; AIDS; infection; cytostatic; anti-HIV;
KW neuroprotective; antiinflammatory; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200172777-A2.
XX
PD 04-OCT-2001.
XX
PF 13-MAR-2001; 2001WO-US08117.
XX
PR 13-MAR-2000; 2000US-0188986.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Hillman JL, Baughn MR, Yue H, Lal P, Lu DAM, Patterson C;
PI Azimzai Y, Bandman O, Tang YT, Mathur P, Shah P, Au-Young J;
PI Reddy R;
XX
DR WPI; 2001-570896/64.
DR P-PSDB; ABB50214.
XX
PT Novel transcription factor polypeptides, used to treat diseases
PT associated with altered activity and expression of TRFX, and to screen
PT for agents capable of modulating its activity -
XX
PS Claim 11; Page 299; 327pp; English.
XX
CC The present sequence is the coding sequence for a human transcription
CC factor. The transcription factor and its coding sequence are useful in
CC the diagnosis, treatment and prevention of diseases associated with
CC altered expression of the transcription factor e.g. cell proliferative,
CC autoimmune/inflammatory, neurological and developmental disorders. A
CC number of specific disorders/diseases are given in the specification,
CC including: arteriosclerosis, cirrhosis, hepatitis, cancers, AIDS,
CC allergies, anaemia, asthma, autoimmune thyroiditis, bronchitis, atopic
CC dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,
CC Grave's disease, multiple sclerosis, osteoarthritis, pancreatitis,
CC psoriasis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative
CC colitis, uveitis, Alzheimer's disease, Huntington's disease, Parkinson's
CC disease, stroke, and viral, bacterial, fungal and protozoal infections.
XX
SQ Sequence 1461 BP; 413 A; 335 C; 298 G; 415 T; 0 other;

Query Match 20.3%; Score 420.8; DB 22; Length 1461;
Best Local Similarity 97.3%; Pred. No. 3.1e-113;
Matches 428; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
Qy   1188 ttcaaattqqqattcqcagaggcatttttggaacatctctgqaaaaaattqcaqqacccca 1247
          || ||  ||||||||||||||||||||||||||||||||| |||||||| |||||||| |||
Db    530 tttaatatgggattcgcagaggcatttttggaacatctttgqaaaaacttgcaggatcca 589

Qy   1248 aqtaatcctgccatcatcaggcaggctgctggaaattatattggaagctttttgqcaaqa 1307
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    590 aqtaatcctgccatcatcaggcaggctgctggaaattatattggaagctttttggcaaqa 649

Qy   1308 qctaaatttattcctcttattactqtaaaatcatgcctagatcttttqqttaactggctg 1367
          |||||||||||| |||||||||||||||| |||||||||||||||||||||||||||||||
Db    650 gctaaatttatttctcttattactgtaaaaccatgcctagatcttttggttaactggctg 709

Qy   1368 cacatatacctaataaccaggattcqqqaacaaaqqcattctqcqatqttqctctccat 1427
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    710 cacatatacctaataaccaggattcqqqaacaaaqqcattctqcqatqttqctctccat 769

Qy   1428 qqaccattttactcaqcctqccaaqctqtqttctacacctttqttttttaqqacacaaqcaa 1487
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    770 ggaccattttactcagcctgccaagctatgttctacacctttqttttttaqqacacaaqcaa 829

Qy   1488 cttttqaqcqqaaacctqaaaqaaqqtttqcaqtatcttcaqaqtctqaattttqaqcqq 1547
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    830 cttttqaqcqqaaacctqaaaqaaqqtttqcaqtatcttcaqaqtctqaattttqaqcqq 889

Qy   1548 ataqtqatqaqccaqctaaatcccctqaaqatttqcctqccctcaqtqqttaacttttt 1607
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    890 ataqtqatqaqccaqctaaatcccctqaaqatttqcctqccctcaqtqqttaacttttt 949

Qy   1608 qctqcaatcacaaataaqta 1627
          ||||||||||||||| | | |
Db    950 qctqcaatcacaaaqatqaa 969
```

RESULT 11
AAX51663
ID AAX51663 standard; cDNA; 437 BP.
XX
AC AAX51663;
XX
DT 21-JUN-1999 (first entry)
XX
DE Human secreted protein 5' EST SEQ ID NO:242.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW ... cream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS Homo sapiens.
XX
PN WO9906549-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01231.
XX
PR 01-AUG-1997; 97US-0905279.
XX
PA (GPST) GENSET.
XX
PI Dujlert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153779/13.
DR P-PSDB; AAY12885.
XX
PT Novel nucleic acids encoding human secreted proteins obtained from
PT cDNA libraries derived from testis, ovary, uterus and spleen tissue
XX
PS Claim 1; Page 344-345; 522pp; English.
XX
CC AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY12681 to
CC AAY12913, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid
CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.

```
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ   Sequence 447 BP; 118 A; 85 C; 102 G; 131 T; 1 other;


  Query Match             20.4%;  Score 420;  DB 20;  Length 447;
  Best Local Similarity   98.4%;  Pred. No. 2.6e-113;
  Matches  423;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      113 cgcttcgtcctctctgcagttaaaaaagctqqqcgcgtcgcaggactqqgatttcaaaatatgcg 172
            ||||||||||||||||| |||| ||||||||||||||||||||||||||||||||||||||||
Db        2 cgcttcgtcctctgcagttaaagacgctgggcgcgtcgcaggactqqgatttcaaaatatgcg  61

Qy      173 tgcattaqaqaaatqactttttcaattctcccccaaqaaaaactqttcqqtttqqtqqaac 232
            |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       62 tgcattaqaqaatqatttttcaattctcccccaaqaaaaactqttcqqtttqqtqqaac 121

Qy      233 tqtqacaqaaqtcttqctqaaqtacaaaaaqqqtqaaacaaatqacttqaqttqttaa 292
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      122 tqtqacaqaaqtcttqctqaaqtacaaaaaqqqtqaaacaaatqacttqaqttqttaa 181

Qy      293 qaacqaqctqttaqatccaqacataaaqqatqacqaqatcatcaactqqctqctqaqaatt 352
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      182 qaacqaqctqttaqatccaqacataaaqqatqacqaqatcatcaactqqctqctqaqaatt 241

Qy      353 cqttcttctatcatqtacttqacaaaaqacttqaqcaacttatcaqtattatattaaq 412
            |||||||||| ||||||||||||||||||||||||||||||||||||||||||| ||
Db      242 cqttcttctqtcatqtacttqacaaaaqacttqaqcaacttatcaqtattatattaaq 301

Qy      413 attqccttqqttqaatqaaqtcaaacaqtaqtqqaaaqaqtatttqqctttttcttqqtaa 472
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      302 attqccttqqttqaatqaaqtcaaacaqtaqtqqaaaqaqtatttqqctttttcttqqtaa 361

Qy      473 tcttqtatcaqcacaqactqtttttcctcaqacctqtctcaqcatqattqctteccattt 532
            |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      362 tcttqtatcaqcaqaqactqtttttcctcaqacctqtctcaqcatqattqctteccattt 421

Qy      533 tqtqcctccc 542
            ||:||||||
Db      422 tqwqcctccc 431


RESULT  12
AAS91991
ID   AAS91991 standard; cDNA; 560 BP.
XX
AC   AAS91991;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #27795.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE ) HYSEQ INC.
XX
PI   Dunnwald ET, Lin C, Tang YT;
XX
DR   WPI; 2001-649862/73.
DR   P-PSDB; AB627804.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 27795; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantifating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 560 BP; 178 A; 96 C; 109 G; 177 T; 0 other;


  Query Match             19.2%;  Score 398;  DB 23;  Length 560;
  Best Local Similarity   89.0%;  Pred. No. 9.4e-107;
  Matches  486;  Conservative    0;  Mismatches   55;  Indels    5;  Gaps    5;

Qy      202 cccccaaaaaaaactqttcaqtttqqtqaaactqtqacaqaaqtcttqctqaaaqtacaaaa 261
            ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        1 cccccaqaaaaactqttcaqtttqqtqaaactqtqacaqaaqtcttqctqaaqtacaaaa  60

Qy      262 aqqqtqaaacaaatqacttqaqttqttqaaqaaccaqctqttaqatccaqacataaaqq 321
            |||||||||||||||||||||||||||||||||||||||||||||| ||||| ||
Db       61 cqqqtqaaacaaatqacttqaqttqttqaaqaaccaqctqttaqatccaqccataaqqq 120

Qy      322 atqa-ccaqatcatcaactqqctqctaqaattccqttcttctatcatqtacttqa caaa 379
            |||| |||||||||||||||||||||||||||||||||||| |||||||||||| ||||
Db      121 atqaccaqatcatcaactqqctqctaqaattccqttcttctqtcatqtacttqaccaaa 180

Qy      380 aqacttqaq-caacttatcaqtattatattaaqattqccttqqttqaataqaaqtc aa 437
            |||||||||| |||||||||||||||||| |||||||||||||||||||||| ||
Db      181 aqacttqaqccaacttatcaqtattatattaaqattqccttqqttqaataqaaqtcaaa 240

Qy      438 acaqtaqtqqaaqaqtatttqqctttcttqqtaatcttqtatcaqccacaaactqtttc 497
            |||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db      241 acaqtaqtqqaaqaqtatttqqctttcttqqtaatcttqtatcaqccataaactqtttc 300

Qy      498 ctcaqaccqtqtctcaqcatqattqctt-cccatttqtqcctccccqaqtqatcatta 556
            |||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db      301 ctcaqaccqtqtctcaqcatqattqcttcccccatttqtqcctccccqaqtqatcatta 360

Qy      557 qqaaqqcqatqtaqatqtttcaqattctqatqatqaaqatqataatcttcctqcaaattt 616
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 qqaaqqcqatqtaqatqtttcaqattctqatqatqaaqatqataatcttcctqcaaattt 420

Qy      617 tqacacatqtcacaqaqccttqcaaatataqccaqatatqtaccatcqacaccatqttt 676
            |||||||| |||||||||||||||||||||||||||||||||||  |  |  ||
Db      421 tqacacatatcacaqaqccttqcaaatataqccaqatatqtaccatcatqaqtatactt 480

Qy      677 tctcatqccaatactqqtqqaaaaattccattttqttcqaaaatcaqaaaaacactqqa 736
            | | |     | || | |         | | |   ||| || | ||   | |
```

```
Db    481 ttccttattttgaatgtttaattctcaaqaaaattgtaatcaattaqtaaaaattataaa 540

Qy    737 atqtta 742
          ||||||
Db    541 atgtta 546


RESULT  13
AAS92254
ID   AAS92254 standard; cDNA; 3175 BP.
XX
AC   AAS92254;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #28058.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG28067.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 28058; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3175 BP; 811 A; 849 C; 779 G; 736 T; 0 other;


  Query Match             18.9%;  Score 390.6;  DB 23;  Length 3175;
  Best Local Similarity   63.9%;  Pred. No. 4e-104;
  Matches  927;  Conservative    0;  Mismatches  219;  Indels  305;  Gaps   10;

Qy     69 atqqcqqcaccqctqcttcacacqcqtttqccqqqaqatqcqqccqcttcqtccctctqca 128
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 atqqcqqcaccqctqcttcacacqcqtttqccqqqaqatqcqqccqcttcqccctctqca 60

Qy    129 qttaaqaaqctqqqcqcqtcqaqqactqqqatttcaaatatqcqtqcattaqaaqaatqac 188
          || |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 qtcaaqatqctqqqcqcqtcqaqqactqqqatttcaaatatqcqtqcattaqaaqaatqat 120

Qy    189 tttttcaattctcccccaaqaaaaactqttcqqtttqqtqqaactqtqacaqaaqtcttq 248
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    121 tttttcaattctcccccaaqaaaaactqttcaqtttqqtqqaactqtqacaqaaqtcttq 180

Qy    249 ctqaaqtacaaaaaqqqtqaaacaaatqacttqaqttqttqaaqaaccaqctqttaqat 308
          |||||||||+|| ||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ctqaaqtacaaaacqqqtqaaacaaatqacttqaqttqttqaaqaaccaqctqttaaat 240

Qy    309 ccaqacataaaqqatqaccaqatcatcaactqqctqctaqaattccqttcttctatcatq 368
          |||||||||||||| |       | ||    | || |                  ||
Db    241 ccaqacataaaqqtqqtqtqcccqccatqccqatqatcctqaqqqaqqqqaqqccatt 300

Qy    369 tacttqacaaaaqactttqaqcaacttatcaqtattatattaaqattqccttqqttqaat 428
                  | |||| | |  |||    |  ||   ||  | |       |
Db    301 gtqqqcqaqtqtqccttttcatcccctttcqtqqctttcaqttqtqctcaaaccaqaqtqqc 360

Qy    429 aqaaqtcaaacaqtaqtqaaaqaqtatttqqctttttcttqqtaatcttqtatatcaqcacaq 488
            || |  |  || |   |   |                    | ||     |  | |
Db    361 ctcaqcccqaaqqttqctaaqctq----------  ----cctqcqqtqqqtccq 399

Qy    489 actqttttcctcaqaccqtqtctcaqcatqattqcttcccatttttqtqcctcccccaqttq 548
           | ||   |  | | ||      ||||   |     | |||       |  | || || |
Db    400 cccqtqqcctqcccqcccacatacaqcccccaqcctctqccacaaccccccaaccqcccaqcq 459

Qy    549 atcattaaqqaaqqcqatqtaqatqtttcaqattctqatqatqaaqatqataatcttcct 608
            ||  ||    |      | | |
Db    460 caqcttctqqctcactcatctqctttqccq---------------------------- 489

Qy    609 qcaaattttqacacatqtcacaqaqccttqcaaataataqccaqqatatqtaccatcqaca 668
                                                                  ||||
Db    490 ------------------------------------------------------ qaca 493

Qy    669 ccqtqqtttctcatqccaatactqqtqqaaaaaattttccatttqttcqaaaatcaqaqaqa 728
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    494 ccqtqqtttctcatqccaatactqqtqqaaaaaattttccatttqttcqaaaatcaqaqaqa 553

Qy    729 acactqqaaatqttacqttcataacttactaaqqattaqtqtatatttttccaaccttqaaq 788
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    554 acactqqaatqttacqttcataacttactaaqqattaqtqtatatttttccaaccttqaaq 613

Qy    789 catqaaattctqqaqcttattattqaaaaactactcaaqttqqatqtqaatqcatccccaq 848
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    614 catqaaattctqqaqcttattattqaaaaactactcaaqttqqatqtqaatqcatccccaq 673

Qy    849 caqqqtattqaaqatqctqaaqaaacaqcaactcaaacttqtqqtqqqacaqattccacq 908
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    674 caqqqtattqaaqatqctqaaqaaacaqcaaatcaaacttqtqqtqqqacaqattccacq 733

Qy    909 gaaqqattqtttaatatqqatqaaqatqaaqaaactqaacatqaaacaaaqqctqqtcct 968
          |||||||||| ||||||||
Db    734 qaaqqattqtqtaatatqq---------------------------              752

Qy    969 gaacqqctcqaccaqatqqtqcatcctqtaqccqaqcqcctqqacatcctqatqtctttq 1028

Db    753 ------------------------------------------------------------ 752

Qy   1029 gtttqtctctacatqaaqqatqtctqctatqtaqatqqtaaqqttqatacqqcqaaaca 1088
            || |   | || ||  ||      |  ||| || |  | |   | | |  |  ||||
Db    753 --ttaqcaqtttattaatqaaaqtqqaqatqaaqtttatcataatcaaaqqqtqqaaaca 810

Qy   1089 aqqatctatatcqcqacctqataaacatctttqacaaactcctqttqcccaccatqcc 1148

Db    811 q---------------------------------------------------------- 811
```

```
Qy       1149 tcctgccatgtacagtttttcatgttttacctctgtagtttcaaattgggattcgcagag 1208
                                            ||          |||   ||  |||||||||||
Db        812                              ---ctagtgctgctcatctttggattcgcagag 841

Qy       1209 gcattttggaacatctctggaaaaaattgca-ggacccaagtaatcctgccatcatcag 1267
              |||||||||||| ||| |||||||| |||| ||| ||||||||||||||||||||||||
Db        842 gcattttggaacctctttggaaaaaactgcagggatccaagtaatcctgccatcatcag 901

Qy       1268 gcaggctgctggaaattatattggaagcttttggcaagagctaaattattcctcttat 1327
              |||||||||||| |||||||| |||||||||||||||||||||| ||||||| ||||||
Db        902 gcaggctgctggacattatattgcaagcttttggcaagagctacatttatttctcttat 961

Qy       1328 tactgtaaaatcatgcctagatcttttggttaactggctgcacatataccttaataacca 1387
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 tactgtaaaaccatgcctagatcttttggttaactggctgcacatataccttaataacca 1021

Qy       1388 ggatt cggggaacaaaggcattctgcgatgttg-ctctccatgga ccattttactcagc 1444
              ||||| |||||||||||||||||||||||||||| |||||||||| |||||||||||||
Db       1022 ggattccggggaacaaaggcattctgcgatgttgcctctccatggacccattttactcagc 1081

Qy       1445 ctg ccaagctgtgttctacacctttgtttttag-acacaagcagcttttgagcggaaac 1502
              ||| ||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1082 ctgcccaagctgtgttctacacctttgttttagaacacaagcagcttttgagcggaaac 1141

Qy       1503 ctgaaagaagg 1513
              |||||||||||
Db       1142 ctgaaagaagg 1152
```

RESULT 14
AAS92248/c
ID AAS92248 standard; cDNA; 585 BP.
XX
AC AAS92248;
XX
DT 13 FEB 2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #28052.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067 A2.
XX
PD 11 OCT 2001.
XX
PF 30 MAR 2001; 2001WO-US08631.
XX
PR 31 MAR 2000; 2000US-0540217.
PR 23 AUG 2000; 2000US 0649167.
XX
PA (HYSE) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P PSDB; ABG28061.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 1; SEQ ID No 28052; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct sequences.
XX
SQ Sequence 585 BP; 148 A; 152 C; 123 G; 162 T; 0 other;

```
  Query Match             18.7%;  Score 387;  DB 23;  Length 585;
  Best Local Similarity   97.3%;  Pred. No. 1.7e-103;
  Matches  404;  Conservative    0;  Mismatches   10;  Indels    1;  Gaps    1;

Qy          4 gaggctgtggctggaaggagctgggcatccggcctgaggcgcagcggtcgcgttagttcg 63
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        453 GAGGCTGTGGCTGGAAGGAGCTGGGCATCCGGCCTGAGGCGCAGCGGTCGCGTTAGTTCG 394

Qy         64 gcccaatggcggcaccgctgcttcacacg cgtttgccgggagatgcggccgcttcgtcc 122
              ||||||||||||||||||||||||||||   ||||| ||||||||||||||||||||||
Db        393 GCCCAATGGCGGCACCGCTGCTTCACACGTTGTTTGTCGGGAGATGCGGCCGCTTCGTCC 334

Qy        123 tctgcagttaagaagctgggcgcgtcgaggactgggatttcaaatatgcgtgcattagag 182
              |||||||| |||| |||||||||||||||||||||||||||||||||||||||||||||
Db        333 TCTGCAGTCAAGACGCTGGGCGCGTCGAGGACTGGGATTTCAAATATGCGTGCATTAGAG 274

Qy        183 aatgactttttcaattctcccccaagaaaaactgttcggtttggtggaactgtgacagaa 242
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        273 AATGATTTTTTCAATTCTCCCCCAAGAAAAACTGTTCGGTTTGGTGGAACTGTGACAGAA 214

Qy        243 gtcttgctgaagtacaaaaagggtgaaacaaatgactttgagttgttgaagaaccagctg 302
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 GTCTTGCTGAAGTACAAAAAGGGTGAAACAAATGACTTTGAGTTGTTGAAGAACCAGCTG 154

Qy        303 ttagatccagacataaaggatgaccagatcatcaactggctgctagaattccgttcttct 362
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        153 TTAGATCCAGACATAAAGGATGACCAGATCATCAACTGGCTGCTAGAATTCCGTTCTTCT 94

Qy        363 atcatgtacttgacaaaagactttgagcaacttatcagtattatattaagattgc 417
              ||||||||||||||||||||||||||||||||||||||||||||| |   || |
Db         93 GTCATGTACTTGACAAAAGACTTTGAGCAACTTATCAGTATTATATTGATCTTCC 39
```

RESULT 15
AAS87636
ID AAS87636 standard; cDNA; 1017 BP.
XX
AC AAS87636;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #23440.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX

PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG23449.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 23440; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1017 BP; 301 A; 214 C; 230 G; 272 T; 0 other;

Query Match 15.2%; Score 315; DB 23; Length 1017;
Best Local Similarity 91.7%; Pred. No. 4e-82;
Matches 333; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

```
Qy   1188 ttcaaattqqqattcqcaqaqqcatttttqqaacatctctqqaaaaaattqcaggaccca 1247
          || ||  |||||||||||||||||||||||||||||||| |||||||| |||||||| |||
Db    445 tttaatatgggattcgcagaggcatttttggaacatctttggaaaaacttgcaggatcca 504

Qy   1248 agtaatcctgccatcatcaggcaggctgctggaaattatattgqaaaqctttttqqcaaqa 1307
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    505 agtaatcctgccatcatcaggcaggctgctggaaattatattggaaagctttttggcaaga 564

Qy   1308 gctaaatttattcctcttattactgtaaaatcatgcctagatcttttgqttaactggctg 1367
          |||||||||||| |||||||||||||||| ||||||||||||||||||||||||||||||
Db    565 gctaaatttatttctcttattactgtaaaaccatgcctagatcttttggttaactggctg 624

Qy   1368 cacatatacattaataaccaggattcgggaacaaaggcattctgcgatgttgctctccat 1427
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    625 cacatatacattaataaccaggattcgggaacaaaggcattctgcgatgttgctctccat 684

Qy   1428 ggaccattttactcagcctgccaagctgtgttctacacctttgtttttagacacaagcag 1487
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    685 ggaccattttactcagcctgccaagctgtgttctacacctttgtttttagacacaagcag 744

Qy   1488 cttttgagcggaaacctgaaagaaggtttgcagtatcttcagagtctgaattttgagcgg 1547
          |||||||||||||||||||||||  |  |         ||  ||||  |  |||  ||
Db    745 cttttgagcggaaacctgaaagaaagtctccctcatgacagagcatctggctggtgacggg 804

Qy   1548 ata 1550
          | |
Db    805 aaa 807
```

Search completed: July 30, 2002, 03:59:40
Job time: 4839 sec